Universidade do Minho
Escola de Ciências

Julio César Chávez Galarza

Population genomics and landscape genetics of the Iberian honey bee (*Apis mellifera iberiensis*)

junho de 2016

Julio César Chávez Galarza **Population genomics and landscape genetics of the Iberian honey bee (*Apis mellifera iberiensis*)**

UMinho | 2016

Universidade do Minho
Escola de Ciências

Julio César Chávez Galarza

Population genomics and landscape genetics of the Iberian honey bee (*Apis mellifera iberiensis*)

Tese de Doutoramento em Biologia Molecular e Ambiental
Especialidade em Evolução, Biodiversidade e Ecologia

Trabalho efetuado sob a orientação da
Professora Doutora Maria Alice da Silva Pinto
e coorientação do
Professor Doutor Filipe José Oliveira Costa
e do
Professor Doutor John Spencer Johnston

junho de 2016

DECLARAÇÃO DE INTEGRIDADE

Declaro ter atuado com integridade na elaboração da presente tese. Confirmo que em todo o trabalho conducente à sua elaboração não recorri à prática de plágio ou a qualquer forma de falsificação de resultados.

Mais declaro que tomei conhecimento integral do Código de Conduta Ética da Universidade do Minho.

Universidade do Minho, 14 de Junho de 2016

Nome completo: JULIO CÉSAR CHÁVEZ GALARZA

Assinatura: 

Dissertação apresentada à Escola de Ciências da Universidade do Minho para obtenção do grau de Doutor em Biologia Molecular e Ambiental.

Este trabalho foi realizado no Centro de Investigação de Montanha do Instituto Politécnico de Bragança, sob a orientação da Professora Doutora Maria Alice da Silva Pinto, no Centro de Biologia Molecular e Ambiental da Universidade do Minho, sob co-orientação Professor Doutor Filipe José Oliveira Costa, e no Departamento de Entomologia da Texas A & M University, sob co-orientação do Professor Doutor John Spencer Johnston.

A sua execução foi financiada pela Bolsa de Doutoramento SFRH/BD/68682/2010, atribuída pela Fundação para a Ciência e a Tecnologia (FCT).

Ao Jeová Deus por tudo

Aos meus pais Julio y Claudina pelo
seu amor e apoio em todo momento

Á Yolanda e a Sofia pelo amor,
paciência e apoio durante este desafio

À memória do meu irmão Carlos Rommel Chávez Pineda

AGRADECIMENTOS

Manifesto de maneira muito especial o meu agradecimento à Prof. Doutora Maria Alice da Silva Pinto, orientadora deste trabalho, por todo apoio, dedicação, determinação, paciência, constante estímulo, e em particular por todos os ensinamentos que me transmitiu.

Ao Prof. Felipe Costa, pelo seu apoio incondicional na realização desta tese.

Ao Prof. John Spencer Johnston da Texas A&M University, nos Estados Unidos, pela co-orientação e valiosas sugestões na realização desta tese.

Ao Prof. Lionel Garnery do Laboratoire de Evolution, Génomes et Spéciation, CNRS, na França, pela colaboração, disponibilidade e apoio científico na análise mitocondrial.

Ao Prof. José Rufino da Escola Superior de Tecnologia e Gestão, do Instituto Politécnico de Bragança pelo apoio no suporte computacional e análises de softwares.

Ao Dr. Miguel Carneiro do CIBIO, Universidade do Porto, pelo auxílio e conselhos nas análises de cline geográficos.

Aos professores do laboratório de informação geográfica Dr. João Carlos Azevedo e Dr. João Paulo Castro pela ajuda durante a aprendizagem dos programas SIG.

Aos Professores. Doutores Célia Pais, Cândida Lucas, Rui Oliveira e Ana Paula Sampaio da Escola de Ciências da Universidade do Minho pelo seu apoio, confiança, conselhos e amizade.

À Yolanda e a Sofia pela compreensão, companhia, esperança, felicidade e amor, que se voltaram para mim em símbolos de perseverança e ânimos para seguir em frente.

Aos colegas do Instituto Politécnico de Bragança: Ângelo e Fernando pela amizade e longas tertúlias, ao Andrés pela amizade, confiança e companhia, ao Eric e Marisa pelo apoio, confiança e confidências, ao Miguel por dar exemplo sinceridade y valores, e em especial para a Dora e Cátia

por me aturar meu feitio particular e que se voltaram num apoio fundamental de amizade, confiança, e profissionalismo científico para a conclusão desta tese.

Aos meus muito estimados amigos Andreia e Flávio pelo apoio, cuidado e companhia que brindaram à Yolanda durante o difícil e bonito processo de ser mãe, e aliviaram a minha ausência durante o desenvolvimento desta tese.

Ao Damien e a Ophélie do Laboratoire de Evolution, Génomes et Spéciation, CNRS, na França, pelo recebimento e apoio durante a minha estadia na França

À família Alves Pinto Pacheco, pela hospitalidade brindada aminha pessoa e família desde a minha chegada a Portugal.

Aos professores da Universidade Nacional de Trujillo-Perú: Dr. Manuel Fernández Honores e Ms. C. Elmer Alvítez Izquierdo por me terem apoiado e ensinado grandes valores. Um agradecimento muito especial ao Professor Roberto Rodríguez Rodríguez por ter sido o meu guia durante a vida universitária e modelo de pessoa.

Aos meus irmãos Gianmarco, Michael, Lizbeth, Pilar, Emilia, Ruth, Maira e Milagros, pelo apoio incondicional.

Ao meu tio Miguel pela sua confiança e apoio incondicional desde o início dos meus estudos.

Aos meus tios Ramón, Picolín, Doris, Luz e Carlos e primos pelo seu apoio e confiança que sempre têm posto em mim.

À família Rodríguez Gonzales: Prof. Roberto, Sra. Rocio, Mireya, Lalo e Brandon, pelo apoio e amizade nos momentos mais difíceis. Às famílias Salgado Rodríguez, Juárez Pita, Rebaza Rodríguez, Massé Miguel, sempre me fizeram sentir como parte delas.

Ao Prof. Angel Castro e à primeira turma da Escola de Biotecnologia da Universidade Nacional do Santa pela oportunidade e confiança que me ofereceram para ser professor dessa instituição.

ABSTRACT

The goal of this study was to disentangle the complex variation patterns of the Iberian honey bee (*Apis mellifera iberiensis*) hybrid zone using the highly polymorphic tRNA^{leu}-cox2 mitochondrial region and nuclear genome-wide Single Nucleotide Polymorphisms (SNPs). Initially, a maternal analysis was performed using a PCR-RFLP marker, known as the *Dra*I test, in the tRNA^{leu}-cox2 intergenic region of colonies sampled in Portugal (N=950). Using this test, 16 novel haplotypes of African ancestry, 15 belonging to sub-lineage A_{III} and 1 to sub-lineage A_{I} were identified. This result suggests that the Atlantic side of the Iberian Peninsula is an important reservoir of maternal diversity that has been missed out because of under-sampling in previous studies.

To obtain a fuller picture of maternal diversity patterns in the Iberian honey bee, 711 drones, sampled across three north-south Iberian transects, were screened for tRNA^{leu}-cox2 variation using sequence data. The tRNA^{leu}-cox2 sequence revealed a more complex diversity pattern of haplotypes in the Iberian honey bee than previously thought when using only the *Dra*I test, which is reflected by detection of 164 novel haplotypes of African (lineage A) and Western European (lineage M) ancestry. At the same time, the distribution of haplotypes A and M reported in this study has further refined the well-defined Southwestern-Northeastern clinal pattern previously described, and also has rescued and reinforced the hypothesis of a hybrid origin for the Iberian honey bee. The distribution pattern of both lineages suggests the presence of two glacial refuges located in the Northeastern and Southern of Iberia. Nonetheless, the confined distribution of sub-lineage A_{III} to the North Atlantic side of Iberia suggests a putative third refuge, an hypothesis that deserves further investigation. A phylogenetic tree representing over 281 haplotypes, of which 8 exhibited intermediate A and M characteristics, resolved that the A_{I}, A_{II}, A_{III} cluster and lineage M are sister groups, which have probably diverged from a common ancestor similar to sub-lineage Z. These findings suggest that lineage M might have a more ancestral African origin, not from North African populations but from Northeastern African and Near East populations where haplotypes of sub-lineage Z have been reported. The low number of colonies (only 1) carrying haplotypes of Eastern European ancestry (lineage C) indicates that importation of commercial honey bees is residual in Iberia, which contrasts with other regions of Western Europe where C-lineage honey bees are replacing the native subspecies.

Analysis of selection was carried out to find out if this evolutionary force has been shaping the complex patterns of Iberian honey bees. The 711 drones were screened for 1536 SNPs using

the GoldenGate Assay of Illumina, of which only 383 revealed polymorphic. While a total of 69 outlier SNPs were identified using two coalescent and two Bayesian methods, only 17 were cross-validated by the four methods. Additionally, a spatial method identified 33 outlier SNPs of which 28 coincided with the previous four methods. Among the 17 outliers, 10 exhibited the strongest signal of selection (9 directional and 1 balancing). 71 outlier SNPs were located in or near genes involved in putative functions as signaling, structural, metabolism, regulation, transport, and immunity. Vision, xenobiotic detoxification and immune response mechanisms were well represented by several genes under selection. For the vision mechanism, five genes are responsible for neural development, synapse formation, axon guidance, regeneration and production of chromophore (formation of rhodopsin). For xenobiotic detoxification, three genes are responsible for tolerance to plant toxins and insecticide resistance. Finally, the two genes related with innate immune response participate in the phagocytosis process. The spatial analysis corroborated that selection is an underlying force shaping a latitudinal and longitudinal gradient.

Population structure was inferred from both maternal ($tRNA^{leu}$-cox2 intergenic region) and biparental (SNPs) markers using spatial and non-spatial methods. The geographical distribution of lineages A and M revealed a sharp southwestern-northeastern maternal cline. Congruent with the mtDNA, population structure inferred from SNPs indicates the presence of two clusters that remarkably overlap the distribution of A and M maternal lineages. These results support a process of secondary contact between divergent honey bee populations in Iberia. Further support for the secondary contact hypothesis comes from (i) geographic cline analysis, which revealed the existence of multiple coincident clines, (ii) elevated values of linkage disequilibrium and of diversity towards the center of the cline. The previous findings suggest the existence of two putative refuges located in Northeastern (between Iberian Mountain Range and Pyrenees) and Southern (Betic Ranges in Spain) Iberia, as proposed for a growing list of other Iberian taxa.

In summary, the complex diversity patterns exhibited by Iberian honey bees has seemingly been shaped by a process of secondary contact between two highly divergent groups (A and M), together with selective forces producing local adaptation to a very heterogeneous area as the Iberian Peninsula.

RESUMO

O objectivo deste estudo é a compreensão dos padrões de variação da abelha ibérica (*Apis mellifera iberiensis*), usando uma região mitocondrial altamente polimórfica e os polimorfismos de nucleótido únicos (SNPs) localizados ao longo do genoma nuclear. Inicialmente foi efectuada uma análise dos marcadores maternos em colónias amostradas em Portugal (N=950) para isso usou-se PCR-RFLP, conhecido por teste *Dra*I, na região intergénica tRNA^{leu}-cox2. Usando este teste, 16 novos haplótipos foram identificados com ancestralidade Africana, 15 pertencentes à sub-linhagem A_{III} e 1 à sub-linhagem A_{I}. Estes resultados sugerem que o lado Atlântico da Península Ibérica é um reservatório importante da diversidade materna, que ficou negligenciado devido à amostragem efectuada pelos estudos anteriores não incorporar esta região geográfica.

Para obter um panorama completo dos padrões de diversidade materna da abelha ibérica, 711 machos amostrados em três transectos norte-sul ao longo da Península Ibérica, foram estudados usando a variação da sequência da região intergénica tRNA^{leu}-cox2. A sequência tRNA^{leu}-cox2 revelou um padrão de diversidade de haplótipos mais complexo do que o que anteriormente se tinha verificado quando se usou apenas o teste *Dra*I, aqui observou-se a presença de 164 haplótipos novos de ancestralidade Africana (linhagem A) e da Europa Ocidental (linhagem M). Ao mesmo tempo, a distribuição dos haplótipos A e M reportado neste estudo refinou o cline bem definido Sudoeste-Nordeste, o que resgatou e reforçou a hipótese de uma origem híbrida das abelhas ibéricas. A distribuição do padrão das duas linhagens sugere a presença de dois refúgios glaciares localizados no Nordeste e Sul da Ibéria. Não obstante, a distribuição confinada da sub-linhagem A_{III} no lado Atlântico Norte da Ibéria sugere um terceiro refúgio potencial, uma hipótese que necessita de estudos adicionais. A árvore filogenética onde se encontram mais de 281 haplótipos, 8 dos quais mostram características intermédias entre a linhagem A e M, mostra que o grupo A_{I}, A_{II}, A_{III} e a linhagem M são grupos irmãos, que provavelmente divergiram de um ancestral comum com características similares à sub-linhagem Z. Estes dados sugerem que a linhagem M pode ter uma origem Africana mais ancestral, não do Norte de África mas sim de populações do Nordeste Africano e Médio Oriente onde os haplótipos de sub-linhagem Z têm sido reportados. O baixo número de colónias (apenas uma) que mostram haplótipos de ancestralidade do leste Europeu (linhagem C) indicam que a importação de abelhas comerciais na Península Ibérica é residual, comparado com outras regiões da Europa Ocidental, onde as abelhas de linhagem C têm vindo a substituir as abelhas locais.

Foram efectuadas análises de selecção para perceber se esta força evolutiva tem vindo a moldar os padrões complexos da abelha ibérica. Os 711 machos foram genotipados para 1536 SNPs usando GoldenGate Assay da Illumina, dos quais apenas 383 eram polimórficos. O total de 69 SNPs foram detectados usando dois métodos coalescente e dois métodos Bayesianos, apenas 17 foram detectados pelos quatro métodos. Adicionalmente, um método espacial identificou 33 SNPs outliers, 28 dos quais coincidem com os quatro métodos prévios. Entre os 17 outliers, 10 mostram um sinal de selecção muito forte (9 direccionais e 1 balanceadora). 71 SNPs outliers estão localizados dentro ou perto de genes, tendo como funções putativas a sinalização, estrutura, metabolismo, regulação, transporte e imunidade. Mecanismos de visão, destoxificação xenobiótica e resposta imune estiveram bem representados por vários genes que mostram SNPs com sinais de selecção. Para o mecanismo da visão, 5 genes são responsáveis pelo desenvolvimento neuronal, formação de sinapses, orientação de axónios, regeneração e produção do cromóforo (formação da rodopsina). Para a destoxificação xenobiótica, três genes são responsáveis pela tolerância a toxinas de plantas e resistência a insecticida. Finalmente, dois genes relacionados com a resposta imune inata, participam no processo de fagocitose. As análises espaciais corroboram que a selecção é uma força que molda o gradiente latitudinal e longitudinal.

A estrutura populacional foi inferida tanto por marcadores maternos (região intergénica $tRNA^{leu}$-cox2) como por biparentais (SNPs) usando métodos espaciais e não-espaciais. A distribuição geográfica das linhagens A e M revelam um cline materno sudoeste-nordeste bem definido. Concordante com o ADNmt, a estrutura populacional inferida pelos SNPs indica a presença de dois grupos, que sobrepões com a distribuição das linhagens maternas A e M. Estes resultados suportam o processo de contacto secundário entre duas populações divergentes de abelhas na Ibéria. Outra evidencias que suportam a hipótese de contacto secundários advém (i) da análise do clino geográfico, que revela a existência de múltiplos clinos coincidentes, (ii) valores elevados de *linkage disequilibrium* e diversidade em direcção ao centro do clino. Estudos prévios sugerem a existência de refúgios potenciais localizados no Nordeste da Ibéria (entre o Sistema Ibérico e os Pireneus) e o no Sul (Sistema Bético em Espanha), tal como sugerido para um número crescente de taxa ibéricos.

Em suma, os padrões complexos de diversidade apresentados pela abelha ibérica, parecem ter sido moldados por processos de contacto secundário de dois grupos divergentes (A e M), juntamente com as forças selectivas que levaram à adaptação local a uma área muito heterogénea como a Península Ibérica.

RESUMEN

El objetivo de este estudio fue desenmarañar los padrones complejos de variación de la zona híbrida de la abeja ibérica (*Apis mellifera iberiensis*) usando la región mitocondrial altamente polimórfica $tRNA^{leu}$-cox2 y polimorfismos de nucleótido simples (SNPs) nucleares del todo el genoma. Inicialmente, un análisis maternal fue llevado a cabo usando un marcador PCR-RFLP, conocido como el test *Dra*I, en la región intergénica de colonias muestreadas en Portugal (N=950). Usando este test, 16 nuevos haplótipos de origen Africano, 15 pertenecientes al sub-linaje A_{III} y 1 a sub-linaje A_{I} fueron identificados. Este resultado sugiere que el lado Atlántico de la Península Ibérica es un importante reservorio de diversidad maternal que ha sido omitido en previos estudios por causa de un muestreo más amplio.

Para obtener una imagen complete de padrones de diversidad materna en la abeja ibérica, 711 machos, muestreados a través de tres transectos norte-sur a lo largo de la Península Ibérica, fueron analizados para la variación de la secuencia de $tRNA^{leu}$-cox2. La secuencia de esta región presentó una mayor complejidad de padrón de diversidad de haplótipos en la abeja ibérica de lo que se había pensado cuando usando el test *Dra*I, lo cual es reflejado en la detección de 164 nuevos haplótipos de origen Africano (linaje A) y Europeo Occidental (linaje M). Al mismo tiempo, la distribución de haplótipos A y M reportados en este estudio, ha extra refinado el bien definido padrón clinal Suroeste- Nordeste previamente descrito, y también ha rescatado y reforzado la hipótesis del origen híbrido de la abeja ibérica. El padrón de distribución de ambos linajes sugiere la presencia de dos refugios glaciares localizados en el Nordeste y Sur de Iberia. No obstante, la distribución confinada del sub-linaje A_{III} al lado Nord Atlántico de Iberia sugiere un putativo tercer refugio, una hipótesis que merece extra investigación. Un árbol filogenético inferido de 281 haplótipos, donde 8 haplótipos exhibieron características intermedias de los linajes A y M, resolvió que el grupo A_{I}, A_{II}, A_{III} y linaje M son grupos hermanos, los cuales probablemente han divergido de un común ancestro similar al sub-linaje Z. Estos hallazgos sugieren que el linaje M podría tener un origen más antiguo, no de poblaciones Nord Africanas sino de poblaciones Nord-Orientales Africanas y de Medio Oriente donde haplótipos del sub-linaje Z han sido reportados. El bajo número de colonias presentando haplótipos de origen Europeo Oriental (linaje C) indica que la importación de abejas comerciales es residual en Iberia, lo cual contrasta con otras regiones de Europa Occidental donde abejas de linaje C están reemplazando a las subespecies nativas.

Análisis de selección fue realizado para averiguar si esta fuerza evolucionaria ha estado formando los padrones complejos de la abeja ibérica. Los 711 machos fueron analizados para

1536 SNPs usando GoldenGate Assay de Illumina, de los cuales solo 383 revelaron polimorfismo. Mientras un total de 69 SNP outliers fueron identificados usando dos métodos de coalescencia y dos métodos bayesianos, solo 17 fueron por los cuatro métodos. Adicionalmente, un método espacial identificó 33 SNP outliers de los cuales 28 coincidieron con los cuatro métodos anteriores. Entre los 17 outliers, 10 exhibieron las más fuertes señales de selección (9 direccionales y 1 balanceadora). 71 SNPs outliers fueron localizados dentro o cerca de genes involucrados en funciones putativas como señalización, estructural, metabolismo, regulación transporte, e inmunidad. Mecanismos de visión, detoxificación xenobiótica y respuesta inmune estuvieron bien representados por varios genes bajo selección. Para el mecanismo de la visión, cinco genes son responsables para desenvolvimiento neural, formación de sinapsis, orientación del axón, regeneración y producción del cromóforo (formación de rodopsina). Para la detoxificación xenobiótica, tres genes son responsables para la tolerancia a toxinas de plantas y resistencia a insecticidas. Finalmente, los dos genes relacionados con la respuesta inmune innata participan en el proceso de fagocitosis. El análisis espacial corroboró que la selección es una fuerza fundamental formando un gradiente latitudinal y longitudinal.

La estructura poblacional fue inferida tanto del marcador materno (región intergénica $tRNA^{leu}$-cox2) como de biparentales (SNPs) usando métodos espaciales y no espaciales. La distribución geográfica de linajes A y M reveló un marcado clino materno Suroeste- Nordeste. Congruente con el ADNmt, la estructura poblacional inferida de SNPs indica la presencia de dos grupos que extraordinariamente se sobreponen a la distribución de los linajes A y M. Estos resultados soportan un proceso de contacto secundario entre poblaciones divergentes de abejas en Iberia. Otras evidencias que soportan la hipótesis de contacto secundario viene de (i) análisis de clinos geográficos, los cuales revelaron la existencia de multiple clinos coincidentes, (ii) valores elevados de *linkage disequilibrium* y de diversidad hacia el centro del clino. Los hallazgos anteriores sugieren la existencia de dos putativos refugios localizados en el Nordeste (entre las Sistema Ibérico y Pirineos) y el Sur de Iberia (Sistemas Béticos en España), como propuesto para una lista creciente de otros taxa ibéricos.

En resumen, los complejos padrones de diversidad exhibidos por la abeja ibérica ha sido aparentemente formado por un proceso de contacto secundario entre dos grupos altamente divergentes (A y M), junto con fuerzas selectivas produciendo adaptación local en un área muy heterogénea como la Península Ibérica.

RESUME

Le but de cette étude était de comprendre les variations complexes qui façonnent la diversité de l'abeille mellifère ibérique (*Apis mellifera iberiensis*) au sein d'une zone hybride utilisant la région mitochondrial particulièrement polymorphe -tRNA^{leu}-cox2- et le polymorphisme d'un seul nucléotide (SNPs). Une analyse des fragments de restriction (PCR-RFLP) de la région intergénique mitochondriale tRNA^{leu}-cox2, connue sous le nom de test *Dra*I a tout d'abord permis de déterminer l'origine maternelle des colonies échantillonnées au Portugal (N=950). 16 nouveaux haplotypes d'origine africaine ont été identifiés parmi lesquels 15 appartiennent à la sous-lignée évolutive A$_{III}$ et 1 à la sous-lignée évolutive A$_{I}$. Ces résultats suggèrent que la côte Atlantique de la Péninsule Ibérique comporte une importante diversité mitochondriale, ce qui, par manque d'échantillonnage, n'avait pas été décrit dans les études précédentes.

Afin d'obtenir un aperçu plus exhaustif de la diversité mitochondriale présente chez l'abeille ibérique, la région intergénique tRNA^{leu}-cox2 a été séquencée pour 711 faux-bourdons échantillonnés le long de trois gradients Nord-Sud de la Péninsule Ibérique. La détection de 164 nouveaux haplotypes des lignées évolutives Africaines (lignée A) et Ouest-Méditerranéenne (lignée M) révèlent une diversité plus complexe que celle attendue avec la seule utilisation du test *Dra*I. La distribution géographique des haplotypes décrits dans cette étude a permis d'une part, d'affiner le modèle évolutif de migration précédemment admis, décrivant une expansion du Sud-Ouest vers le Nord-Est et soutient d'autre part l'hypothèse d'une origine hybride des abeilles mellifères ibériques. La distribution géographique de chacune de ces lignées suggère la présence de deux refuges glaciaires situés au Nord-Est et au Sud de la Péninsule Ibérique. Cependant, la distribution géographique de l'haplotype A$_{III}$, confiné au Nord de la côte Atlantique de la Péninsule Ibérique tend à décrire un troisième refuge glaciaire, hypothèse qui mérite des recherches supplémentaires. Un arbre phylogénétique représentant 281 haplotypes, parmi lesquels 8 montrent des caractéristiques intermédiaires entre les lignées A et M, permet de classer l'ensemble des sous-lignées (A$_{I}$, A$_{II}$, A$_{III}$) et de la lignée M en groupes frères, et que par conséquent, ces haplotypes dériveraient probablement d'un ancêtre commun ayant les caractéristiques de l'actuelle sous-lignée Z. Ces découvertes suggèrent que la lignée évolutive M, aurait, non pas un ancêtre d'origine Nord-Africaine mais plutôt d'Afrique de l'Est et du Proche Orient, lieu où, est actuellement présente la sous-lignée Z. Le faible nombre de colonies (1 uniquement) portant un haplotype d'Europe de l'Est (lignée C) indique que les importations à but commercial sont résiduelles sur la péninsule Ibérique, contrastant avec

d'autres régions de l'Europe de l'Ouest où les importations occasionnent un remplacement progressif de la sous-espèce native.

Une analyse de sélection a été effectuée afin de déterminer si cette force évolutive a participé à la mise complexe de l'abeille Ibérique. Pour 1536 SNPs (GoldenGate Assay of Illumina) criblés sur les 711 faux-bourdons, seulement 383 se sont montrés être polymorphes. Tandis qu'un total de 69 SNPs on été identifiés utilisant deux méthodes de coalescence et deux méthodes Bayésiennes, seulement 17 sont validés par l'utilisation croisée des 4 méthodes. De plus, une méthode d'analyse spatiale a permis d'isoler 33 SNPs d'intérêt, parmi lesquels 28 coïncident avec les résultats précédents. Parmi les 17 SNPs, 10 témoignent d'un fort signal sélection (9 de type directionnelle et 1 de type balancée). 71 SNPs ont été situés au sein ou à proximité de gènes impliqués dans d'hypothétiques fonctions à rôle structural, de signalisation, métabolique, régulateur, de transport et d'immunité. Les mécanismes de vision, de détoxification xénobiotique et de réponses immunitaires sont clairement représentés par de nombreux gènes sous sélection. Pour le mécanisme de vision, cinq gènes sont responsables du développement neuronal, de la formation synaptique, du guidage axonal, de la régénération et de la production de chromophores (production de rhodopsine). Concernant la détoxification xénobiotique, trois gènes sont responsables de la tolérance aux toxines végétales et de la résistance aux insecticides. Enfin les 2 gènes en relation avec la réponse immunitaire innée participent aux mécanismes de phagocytose. L'analyse spatiale appuie le fait que la sélection est une force sous-jacente impliquée dans la mise en place de gradients latitudinaux et longitudinaux.

La structure de la population à été obtenue à la fois grâce aux marqueurs maternel (région intergénique mitochondriale $tRNA^{leu}$-cox2) et biparentaux (SNPs) sur la base de méthodes spatiales et non spatiales. La distribution géographique des lignées évolutives A et M a révélé l'existence d'un remarquable gradient Sud-Ouest/Nord-Est. En accord avec l'ADNmt, la structure de la population inférée par l'étude des SNPs indique la présence de deux groupes qui coïncident remarquablement à la distribution géographique des lignées évolutives maternelles A et M. Ces résultats supposent un processus de contact secondaire entre populations d'abeilles divergents sur la péninsule Ibérique. Les arguments en faveur de l'hypothèse de la remise en contact secondaire sont (i) les analyses des gradients spatiaux, qui révèlent l'existence de plusieurs gradients cohérents et (ii) des valeurs de déséquilibre de liaison croissantes en direction du centre du gradient. Ces résultats suggèrent l'existence de deux refuges hypothétiques situés au Nord-Est

(entre les monts Ibériques et les Pyrénées) et le Sud (Cordillère Bétique) de la péninsule Ibérique qui est considérée comme une zone d'expansion de nombreux taxons.

En résumé, les profils complexes de diversité observés chez l'abeille Ibérique a, semblerait-il, été façonnée à la fois par un mécanisme de remise en contact secondaire entre deux groupes fortement différenciés (A et M) accompagné de forces de sélection conférant des adaptations locales dans la zone hétérogène que constitue la Péninsule Ibérique.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND ACRONYMS

'ahb': Initials to assign Africanized honey bee genome (traces)-derived SNPs

AICc: Akaike Information Criterion corrected

Ala: Alanine

'AMB': Initials to assign honey bee reference genome-derived SNPs

AMOVA: Analysis of Molecular Variance

Arg: Arginine

Asn: Asparagine

Asp: Aspartic acid

AT: Atlantic Transect

BEEBASE: Bee genomic database

BIC: Bayesian Information Criterion

bp: Base pairs

BYM: Convolution Gaussian prior spatial mixture

c: Centre, distance from sampling location

CI: Confidence Interval

Cld: Cloud cover

cM: Centimorgan

***COX1* gene** or ***COI* gene:** Cytochrome Oxidase subunit I

***COX2* gene** or ***COII* gene:** Cytochrome Oxidase subunit II

CpG: Regions of DNA where a cytosine occurs next to a guanine

***csd* gene:** *Complementary sex determiner* gene

CT: Central Transect

CYP6AS7 protein: Cytochrome P450 6AS7

DAPC: Discriminant Analysis of Principal Components

DIC: Deviance Information Criterion

DNA: Deoxyribonucleic Acid

***Dra*I:** Restriction enzyme from *Deinococcus radiophilus* that recognizes TTT^AAA sites

Dscam: Down syndrome cell-adhesion molecule

'est': Initials to assign honey bee expression sequence tag-derived SNPs

ESTs: Expressed Sequence Tags

FLYBASE: Fruit fly genomic database

F_{ST}: Fixation index of genetic differentiation

G: Likelihood ratio

GenBank: Genetic sequence database

Gln: Glutamine

Gly: Glycine

GPS: Global Positioning System

GSTs: Glutathione-S-Transferases

His: Histidine

Iberian honey bee: *Apis mellifera iberiensis*

indels: Insertion or deletion of bases in the DNA

Ins: Insolation

K: Number of cluster

km: Kilometer

km2: Square kilometer

***K*-nearest neighbours:** Algorithm based on distances and number of samples

kWh: Kilowatt hour

Lat: Latitude

LD: Linkage Disequilibrium

LG: Linkage Groups

Lineage A: *African lineage*

Lineage C: *Eastern Europe lineage*

Lineage M: *Western and Northern European lineage*

Lineage O: *Near East and Asia lineage*

Long: Longitude

LRT: Likelihood Ratio Test

m^2: Square meter

Mb: Megabase

MCMC: Markov Chain Monte Carlo

Mitotype: Mitochondrial haplotype

Moran's *I*: Moran spatial autocorrelation

MT: Mediterranean Transect

mtDNA: mitochondrial DNA

Mya: Million years ago

N_a: Mean number of alleles per locus

NCBI: National Center for Biotechnology Information

N_e: Effective number of alleles

NIMC2: Nimrod C2

nm: Nanometer

N_p: Number of private alleles

°C: Celsius degree

PC1: Principal Component 1

PCA: Principal Component Analysis

PCo: Principal Coordinate Analysis

PCR: Polymerase Chain Reaction

PCR - RFLP: Polymerase Chain Reaction - Restriction Fragment Length Polymorphism

P_{LD} Fisher: Percentage of linkage disequilibrium based on Fisher exact test

***P*max:** Maximum allele frequency at the one end of the cline

***P*min:** Minimum allelic frequency at the one end of the cline

Prec: Precipitation

***P*-value:** It is a statistical probability of Karl Pearson

Q: Matrix of membership proportions

***r*:** Pearson's correlation coefficient

r^2: Measure of linkage disequilibrium

RFLP: Restriction Fragment Length

Rh: Relative humidity

RNA: Ribonucleic Acid

Ser: Serine

SNPs: Single Nucleotide Polymorphisms

sPCA: Spatial Principal Component Analysis

Tmax: Maximum temperature

Tmean: Mean temperature

Tmin: Minimum temperature

***tRNA*^{leu} gene:** *transfer RNA leucine* gene

UGT: UDP-Glycosyltransferases

***uh*:** Unbiased haploid genetic diversity

US: United States

Val: Valine

w: Width, 1/maximum slope

''': Haplotype with five Q elements

'': Haplotype with four Q elements

': Haplotype with three Q elements

|r|: Absolute correlation value

3′ UTR: 3′ Untranslated Region

5′ UTR: 5′ Untranslated Region

δ: Delta, distance between the centre and the tail

ΔF_{st}: Rate of change in the F_{st} values over the values of K.

ΔK: Rate of change in the log-probability values over the values of K.

τ: Tau, slope of the tail

Φ_{PT} values: It is a measure of population genetic differentiation for binary or haploid data. It is analogous to F_{st}.

Ψ: Spatial interaction parameter

LIST OF TABLES

LIST OF FIGURES

Publications in the scope of the thesis

❖ Papers published in international peer-review journals

- Pinto MA, Muñoz I, **Chávez-Galarza J**, De la Rúa P (2013) The Atlantic side of the Iberian Peninsula: a hot-spot of novel African honey bee maternal diversity. **Apidologie**, 43: 663-673.
  DOI: 10.1007/s13592-012-0141-1
  Impact factor 2012: 2.155

- **Chávez-Galarza J**, Henriques D, Johnston JS, Azevedo JC, Patton JC, Muñoz I, De la Rúa P, Pinto MA (2013) Signatures of selection in the Iberian honey bee (*Apis mellifera iberiensis*) revealed by a genome scan analysis of single nucleotide polymorphisms (SNPs). **Molecular Ecology**, 22: 5890-5907.
  DOI: 10.1111/mec.12537
  Impact factor 2013: 5.84

- **Chávez-Galarza J**, Henriques D, Johnston JS, Carneiro M, Rufino J, Patton JC, Pinto MA (2015) Revisiting the Iberian honey bee (*Apis mellifera iberiensis*) contact zone: maternal and genome-wide nuclear variations provide support for secondary contact from historical refugia. **Molecular Ecology**, 24: 2973-2992.
  DOI: 10.1111/mec.13223
  Impact factor 2015: 6.494

❖ Paper *in preparation*

- **Chávez-Galarza J**, Garnery L, Henriques D, Neves CJ, Loucif-Ayad W, Johnston JS, Pinto MA. Revisiting maternal patterns of variation of *Apis mellifera iberiensis*: a revision based on sequence data of the $tRNA^{leu}$-cox2 mitochondrial intergenic region. *In preparation*.

❖ Publications in conference proceedings

- **Chávez-Galarza J**, Henriques D, Muñoz I, De la Rúa P, Azevedo J, Patton JC, Johnston JS, Pinto MA. Un escaneo exploratório del genoma de la abeja ibérica para detectar loci candidatos a selección. Proceedings of the II Iberian Congress of Beekeeping. Guadalajara – Spain. 18 – 20 October 2012.

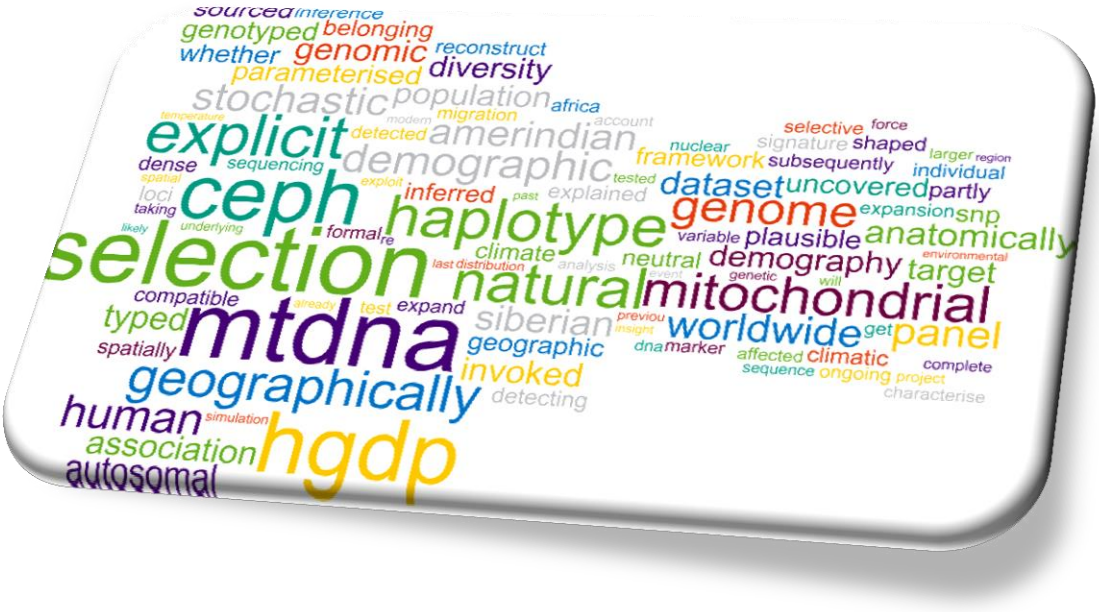
- Henriques D, **Chávez-Galarza J**, Muñoz I, De la Rúa P, Azevedo J, Johnston JS, Pinto MA. Padrões de variação materna da abelha ibérica: um estudo de fina resolução espacial. Proceedings of the II Iberian Congress of Beekeeping. Guadalajara – Spain. 18 – 20 October 2012.

- Pinto MA, Johnston JS, Azevedo J, Muñoz I, **Chávez-Galarza J**, Castro JP, de la Rúa P, Patton JC. Evolutionary history of the Iberian honey bee (*Apis mellifera iberiensis*): A genome-wide approach. Proceedings of the 17th Congress of Management of Common Goods and Sustainable Regional Development and 5th Congress de Management and Nature Conservation. Organized by Portuguese Association for Regional Development and Spanish Association of Regional Science. Bragança – Portugal y Zamora – España. 29 June – 02 July 2011.

❖ Other papers published in international peer-review journals

- Pinto MA, Henriques D, **Chávez-Galarza J**, Kryger P, Garnery L, van der Zee R, Dahle, B, Soland-Reckeweg G, De la Rúa P, Dall'Olio R, Carreck NL, Johnston JS (2014) Genetic integrity of the black honey bee (*Apis mellifera mellifera*) from protected populations: a genome-wide assessment using SNPs and mtDNA sequence data. **Journal of Apicultural Research,** 53(2): 269-278.
  DOI: 10.3896/IBRA.1.53.2.08
  Impact factor 2014: 1.895

- Muñoz I, Henriques D, Johnston JS, **Chávez-Galarza J**, Kryger P, Pinto MA (2015). Reduced SNP panels for genetic identification and introgression analysis in the dark honey bee (*Apis mellifera mellifera*). **PLoS ONE,** 10(4): e0124365.
  DOI: 10.1371/journal.pone.0124365
  Impact factor 2015: 3.23



Chapter I

Motivation, objectives and thesis outlines

Motivation

Iberia is considered a top laboratory for evolutionary studies and the Iberian honey bee (*Apis mellifera iberiensis*) an excellent hybrid zone study model. Hybrid zones are shaped by evolutionary forces acting on the whole genome and on specific regions of chromosomes. Dissection of genome-wide (neutral) from genome-specific effects (adaptive), a crucial goal in evolutionary biology, can be facilitated by combining the tools of population genomics and landscape genetics. This study seeks to promote an integrative approach to the evolutionary history of the Iberian honey bee hybrid zone by examining both mitochondrial and nuclear markers in a landscape genetics context. The tRNAleu-cox2 intergenic mitochondrial region and over 1536 SNP loci were analyzed using a battery of population genomics and landscape genetics tools to identify, spatially represent and interpret patterns of neutral and adaptive variation, and explore the molecular basis of adaptive divergence. This study helped to clarify and reconstruct our view of the Iberian honey bee hybrid zone.

Objectives

The goal of this study was to disentangle the complex variation patterns of the Iberian honey bee hybrid zone. The specific objectives included:

- To uncover, represent, and interpret the spatial mtDNA cline for the entire Iberian Peninsula;
- To identify and interpret the spatial structure generated by loci under selection and potentially reveal the molecular basis of the observed adaptive differentiation in the Iberian honey bee hybrid zone;
- To unravel and interpret the complex Iberian honey bee population structure by comparing patterns of mtDNA and neutral nuclear variations.

Thesis outlines

- **Chapter I** describes the context, motivation and objectives of this thesis, as well as its global structure.
- **Chapter II provides** an overview of the literature related with the theme of the thesis, describing briefly the genus *Apis*, and the evolutionary lineages and geographical distribution of the honey bee (*Apis mellifera*), with special focus on the Iberian honey bee.

- **Chapter III** shows the patterns of maternal variation of Portuguese honey bee populations inferred from the *Dra*I test (PCR-RFLP of the tRNA^{leu}-cox2 intergenic region).
- **Chapter IV** provides a fuller picture of the complex maternal diversity exhibited by the Iberian honey bee. In this chapter, honey bee colonies sampled across Iberia were analyzed for the tRNA^{leu}-cox2 using sequence data and compared with other honey bee subspecies.
- **Chapter V** deals with selection. A search for signals of selection was performed across the Iberian honey bee genome using single nucleotide polymorphisms (SNPs). To decrease the number of false positive outliers a battery of different methodological approaches was employed, taking in to account diverse demographic scenarios.
- **In Chapter VI** the Iberian honey bee population structure was inferred from the tRNA^{leu}-cox2 intergenic region and SNPs. Spatial and non-spatial clustering methods and geographic cline were used to cross-validate the findings, due to the Iberian honey bee complex pattern revealed in previous genetic surveys.
- **Chapter VII** includes the overall conclusions and significance of the work. Suggestions for future work are also exposed.
- **Chapter VIII** includes supplementary data not shown in the other chapters and the pdf versions of the chapters that have been published.

Chapter II

General introduction

An overview to the genus *Apis*

Insect pollinators are necessary for production of one-third of our food as well as for the ecosystem equilibrium. Among the numerous pollinators, the honey bee, *Apis mellifera* L., accomplishes an important function as the most economically valuable pollinator of crops worldwide with its service to agriculture valued at > $200 billion per year worldwide (Le Conte & Navajas 2008; Gallai *et al.* 2009).

The genus *Apis* belongs to a large family of bees, Apidae, which are characterized by the presence of a pollen basket. Taxonomically, this family is classified under the phylum Arthropoda, class Insecta and order Hymenoptera. The earliest phylogeographic studies posited that the center of origin for *Apis* was in Asia, the region in which the greatest number of living species occurs (Ruttner 1988). But, a recent study has suggested an apparent European origin supported by the presence of more fossil *Apis* species dating from Oligocene, spreading into Asia, Africa and North America (Kotthoff *et al.* 2013). Currently, the genus *Apis* includes 12 species validated by morphological and molecular analysis (Engel 1999; Arias & Sheppard 2005). Phylogenetic studies based on nuclear DNA and mitochondrial DNA (mtDNA) markers strongly support the clustering of three distinct groups: cavity-nesting bees (*Apis mellifera*, *A. cerana*, *A. koschevnikovi*, *A. nuluensis, A. breviligula*), giant bees (*A. dorsata*, *A. laboriosa*, *A. binghami*, *A. nigrocincta, A. indica*), and dwarf bees (*A. florea*, *A. andreniformis*) (Arias & Sheppard 2005; Raffiudin & Crozier 2007) (Fig. II - 1). Among these, only *A. mellifera* is widely distributed around the world, with the remaining species restricted to the South-East Asia (Fig. II - 2).

The evolutionary lineage that gave rise to extant *A. mellifera* represents an early split from other cavity-nesting bees, and eventually diversified into subspecies and colonized their present native range in the Middle East, Africa and Europe. Estimates of genetic divergence from mtDNA and nuclear loci suggest that *A. mellifera* and *A. cerana* diverged between 6 and 9 million years ago (Cornuet & Garnery 1991; Arias & Sheppard 2005). The high degree of social organization, which is the most important feature that defines many bee species, evolved independently 28 to 36 Mya ago (Cardinal & Danforth 2011). Bee colonies have a complex colony structure formed by three different castes (queen, worker and drone) specialized to perform distinctive functions. The reproductive biology is variable among the bee species. The honey bee shows polyandry, meaning that a virgin queen can mate with 10 to 15 drones. The number of mattings can vary between species and subspecies (Kerr *et al*. 1962; Adams *et al*., 1977; Hernández-García *et al*. 2009). A

queen can have up to four mating flights (Roberts 1944). Once the queen starts laying eggs, she will never mate again (Winston 1987; Tarpy & Page 2000).

The *complementary sex determiner gene* (*csd* gene) is responsible for the sex determination in the honey bee (Beye *et al.* 2003). Individuals hemizygous (non-fertilized egg) or homozygous (fertilized egg) at this gene will produce drones whereas heterozygous will develop into female queens or workers, a system known as haplodiploidy. The diploid males are sterile and are eaten by workers shortly after emerging (Woyke 1963). The caste differentiation between a queen and a worker results from the selective feeding of the royal jelly protein (Winston 1987).

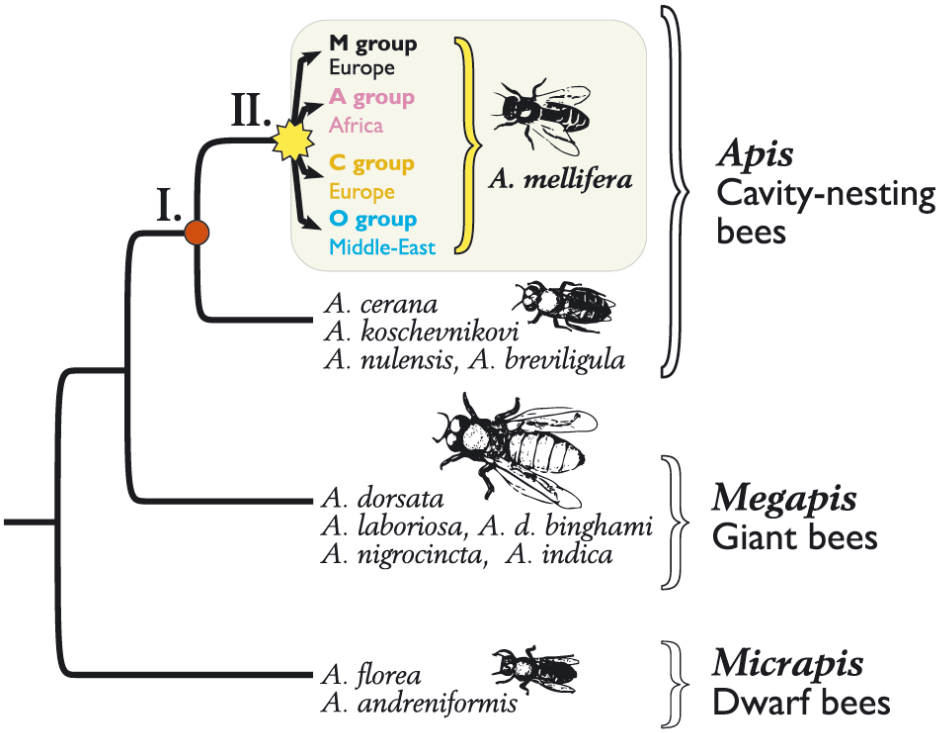


Figure II - 1. Phylogeny representing the three clades of *Apis*. All of the 10 extant *Apis* species apart from *A. mellifera* are found only in Asia. Node I represents the split between *A. mellifera* and other cavity-nesting bees. Node II represents the most recent common ancestor of extant subspecies of *A. mellifera*. (Han *et al.* 2012).

An important milestone in honey bee research was the sequencing of its genome of (Honeybee Genome Sequencing Consortium 2006). The honey bee genome consists of 16 linkage groups with an approximate length of 236 Mbp. The genome is characterized by a high A+T and CpG content, lack of major transposon families, and slower rate of evolution. It shows a great similarity to vertebrates for circadian rhythm, RNA interference and DNA methylation genes. It has fewer genes than *Drosophila* and *Anopheles* for innate immunity, detoxification enzymes, cuticle-

forming proteins and gustatory receptors, but more genes for odorant receptors, and novel genes for nectar and pollen utilization. The number of identified genes was approximately 10000 (Elsik *et al.* 2007), but an update of this prediction has suggested a new gene set of 5000 more protein-coding genes (Elsik *et al.* 2014). Moreover, the honey bee exhibits an extremely high recombination rate of 19 cM/Mb (Beye *et al.*, 2006, Honeybee Genome Sequencing Consortium 2006), which is several-folds higher than that reported for any other higher eukaryotic species.

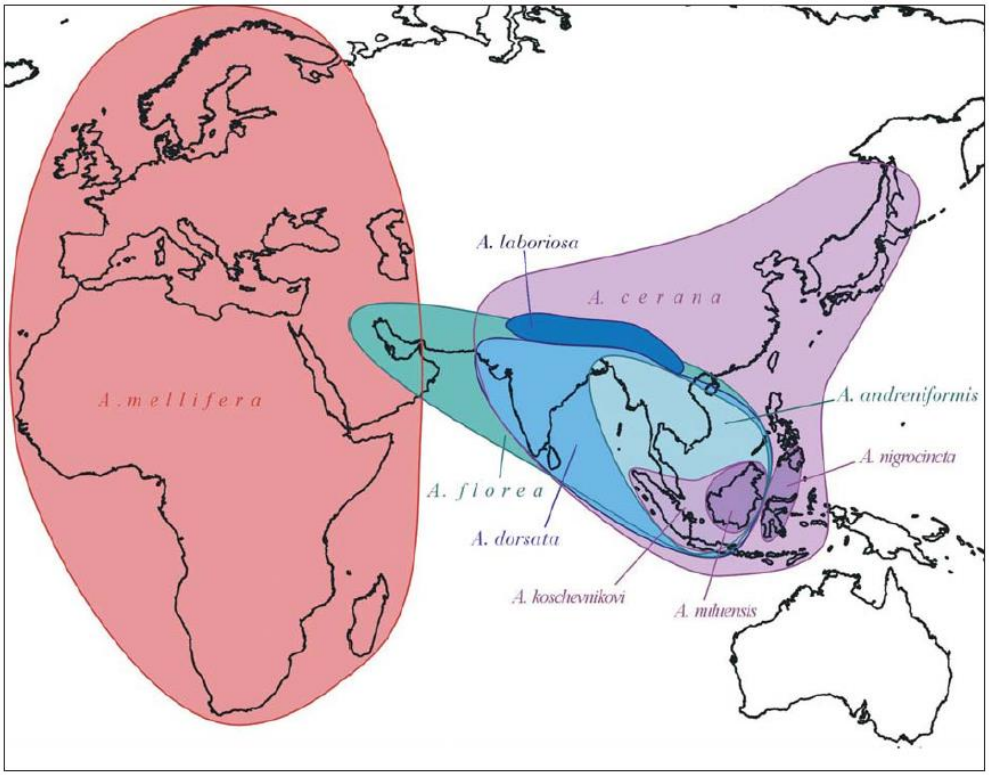


Figure II - 2. Current geographical distribution of *Apis* species (Le Conte & Navajas 2007).

The evolutionary lineages of *Apis mellifera* L.

The honey bee subspecies were first classified according to their morphological and behavioral traits along with their geographical distributions (Ruttner 1988). This analysis grouped 25 subspecies into four evolutionary lineages showing a particular geographic distribution: M lineage in the West Mediterranean and Northwest Europe (*A. m. mellifera*, *A. m. iberiensis*, *A. m. intermissa*, and *A. m. sahariensis*), C lineage in the Southeastern Europe and the Central Mediterranean (*A. m. ligustica, A. m. carnica, A. m. macedonica, A. m. cecropia, A. m. siciliana*), O lineage in the Near East (*A. m. caucasica*, *A. m. anatoliaca*, *A. m. syriaca*, *A. m. meda*, *A. m. armeniaca*, *A. m. cypria*, *A. m.*

adami, A. m. taurica), and A lineage (*A. m. lamarckii*, *A. m. adansonii*, *A. m. scutellata*, *A. m. monticola*, *A. m. litorea*, *A. m. capensis*, *A. m. unicolor, A. m. jemenitica*) in Africa.

With the advent of new molecular techniques and additional morphological studies, the subspecies clustering into the four evolutionary lineages has suffered slight modifications and the number of recognized subspecies has increased to 30 (Ruttner 1988; Engel 1999; Sheppard & Meixner 2003; Meixner *et al*. 2011). The mtDNA variation and new morphological data have indicated that subspecies such as *A. m. intermissa*, *A. m. sahariensis*, and *A. m. siciliana*, as well as the related and later described *A. m. ruttneri* and *A. m. simensis*, must be relocated in lineage A (Cornuet & Garnery 1991; Sheppard *et al.* 1997; Sinacori *et al*. 1998; Franck *et al*. 2000; Franck *et al*. 2001; Kandemir *et al*. 2011; Meixner *et al*., 2011). Only *A. m. mellifera* and *A. m. iberiensis* are included in lineage M, and a new subspecies is described for O lineage: *A. m. pomonella* (Sheppard & Meixner, 2003). The earlier studies of mtDNA did not distinguish between C and O lineages, including both in lineage C (Cornuet & Garnery 1991; Garnery *et al*. 1992). Later, populations from the Near East showed different mitochondrial patterns from the three lineages reported until then (Franck *et al*., 2000; Palmer *et al*., 2000), being associated to morphological O lineage (Franck *et al*. 2000). Similarly, populations belonging to *A. m. jemenitica* exhibited a distinct pattern that was considered a new lineage named Y for Near East (Franck *et al*. 2001). More recent data showed that the previous grouping was wrong, especially for lineage O. These analyses especially made confusion in O lineage. By incorporating new data and by re-analyses, novel mtDNA patterns indicated that lineage O should be reclassified as a sub-lineage of the African lineage and renamed to Z (Alburaki *et al*. 2001). Likewise lineage Y would also be related to lineage A (Meixner *et al*. 2013). Some controversy has arisen regarding to the placement of subspecies in their respective lineage. This is the case of *A. m. cypria*, *A. m. anatoliaca* and *A. m. caucasia* which are morphologically placed in lineage O (Ruttner 1988, Kandemir *et al*. 2011), but the mtDNA analysis suggests that it belongs to lineage C (Garnery *et al*. 1992; Smith *et al*. 1997; Kandemir *et al*. 2006). Additionally, the evolutionary lineage to which the subspecies *A. m. taurica*, *A. m. sossimai* and *A. m. artemisia* belong has not yet been clarified due to lack of information.

Recent analyses based on Single Nucleotides Polymorphisms (SNPs) have also confirmed the presence of the four evolutionary lineages (Whitfield *et al*. 2006; Han *et al*. 2012), as suggested by morphology studies (Ruttner 1988). A more recent study, based on SNPs identified from whole-genomes, has proposed that the four evolutionary lineages diverged approximately 350

000-150 000 years before the present (Middle Pleistocene), and subspecies would have split to 38 000-13 000 years before the present during Late Pleistocene (Wallberg *et al.* 2014).

Three main scenarios have been proposed for the origin of the evolutionary lineages (Fig II-3). The first scenario was proposed by Ruttner *et al.* (1978) using morphological data, which indicated that the origin of lineages was in the Middle East or Northeast Africa, from where they colonized Europe through a western route (via Northwestern Africa to the Iberian Peninsula) and through an eastern route (via Middle East to the Balkans). This hypothesis suggests continuity between lineages A (Africa) and M (Western and Northern Europe) and an ancestral form close to *A. m. syriaca* from Lebanon, Israel, and Jordan. This hypothesis was rescued by Han *et al.* (2012) who reanalyzed the SNP dataset of Whitfield *et al.* (2006) using different measures of genetic distances and tree construction. The second scenario was based mainly on mtDNA analyses (Cornuet & Garnery 1991; Garnery *et al.* 1992). This hypothesis proposes a Middle Eastern origin, but does not suggest a colonization of Europe via a western route. This scenario is based on a phylogenetic tree that groups the lineage A with C rather than with M, arguing against migration across the strait of Gibraltar. Finally, the third scenario was based on 1136 nuclear SNPs (Whitfield *et al.* 2006). This scenario suggests an African origin inferred from a phylogenetic analysis of 14 subspecies of *Apis mellifera* and three outgroups (*A. cerana*, *A. florea*, *A. dorsata*). The phylogenetic root was within lineage A and the colonization of Europe and Asia occurred through an Western expansion of lineage M and one or more eastern expansions of lineages O and C into Europe and Asia, respectively.

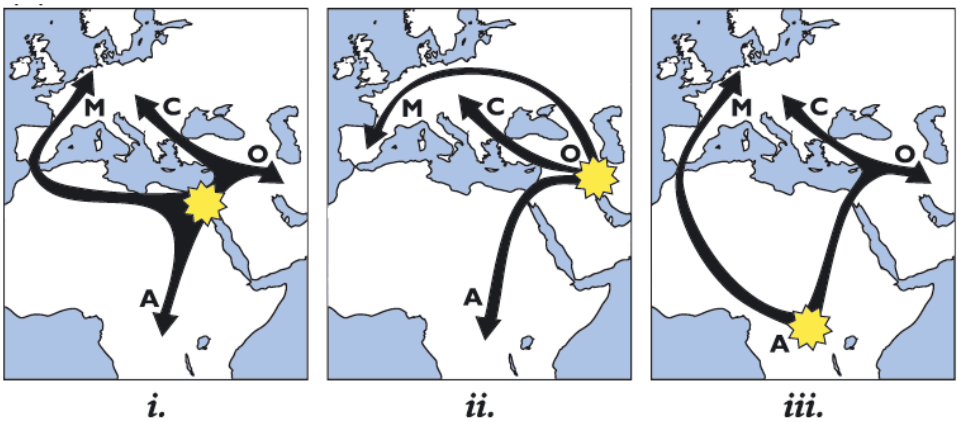


Figure II - 3. Three hypotheses that have been proposed for the origin of *A. mellifera*. (i) An expansion from the Middle East, involving colonization of Europe via two routes, one eastern and one western was first suggested by Ruttner (1978) on the basis of morphometric analyses. (ii) An expansion from the Middle East, which did not involve the western colonization route into Europe was suggested on the basis of trees constructed from mtDNA (Garnery *et al.* 1992). (iii) An origin in Africa was proposed by Wilson (1971) and an expansion out of Africa via both an eastern and western route was suggested by the analysis of >1000 SNPs by Whitfield *et al.* (2006) (Han *et al.*, 2012).

Geographical distribution of *Apis mellifera* L.

The geographical distribution of the 30 honey bee subspecies is described according to Ruttner (1988), Engel (1999), De la Rúa *et al*. (2009), Sheppard & Meixner (2003), Meixner *et al*. (2011) and displayed in Fig. II - 4. The placement of subspecies in its corresponding evolutionary lineage is based on morphological, behavior, allozymes, mtDNA and SNP data (Ruttner 1988; Engel 1999; Garnery *et al*. 1993; Franck *et al*. 2000, 2001; Bouga *et al*. 2005a, b; Kandemir *et al*. 2006; De la Rúa *et al*. 2009; Whitfield *et al*. 2006, Kandemir *et al*. 2011; Han *et al*. 2012).

Western and Northern European lineage (M)

- *Apis mellifera mellifera* Linnaeus (1758), the Western European honey bee or the common European "black bee", is the subspecies with the largest natural area of distribution. It extends from France to Scandinavia and from the British Isles to Central Europe (North of the Alps), including Central Russia.
- *Apis mellifera iberiensis* Engel (1999), the Iberian honey bee, is distributed in the Iberian Peninsula, the Balearic Islands and Macaronesia islands.

African lineage (A)

- *Apis mellifera intermissa* Maa (1953), the Tellian honey bee of the Maghreb, has a tight distribution along the northern coast of Africa as far west as Morocco, into Tunisia in the east, but bordered by the Atlas range in the south.
- *Apis mellifera sahariensis* Baldensperger (1923), the Saharan honey bee, has a tight range in northwestern Africa. It occurs along the southern side of the Atlas range in oases from Morocco to Algeria.
- *Apis mellifera siciliana* Grassi (1880), the Sicilian honey bee, is a subspecies exclusively distributed in the island of Sicily in the Mediterranean Sea.
- *Apis mellifera lamarckii* Cockerell (1906), the Egyptian honey bee, occurs in a narrow range along the Egyptian Nile Valley.
- *Apis mellifera jemenitica* Ruttner (1988), the Arabian or Nubian honey bee, is distributed in countries such as Chad, Oman, Saudi Arabia, Somalia, Sudan and Yemen.
- *Apis mellifera adansonii* Latreille (1804), the West African honey bee, is distributed in Western Africa ranging from Niger in the north , east to Senegal , and as far south as Democratic Republic of Congo.

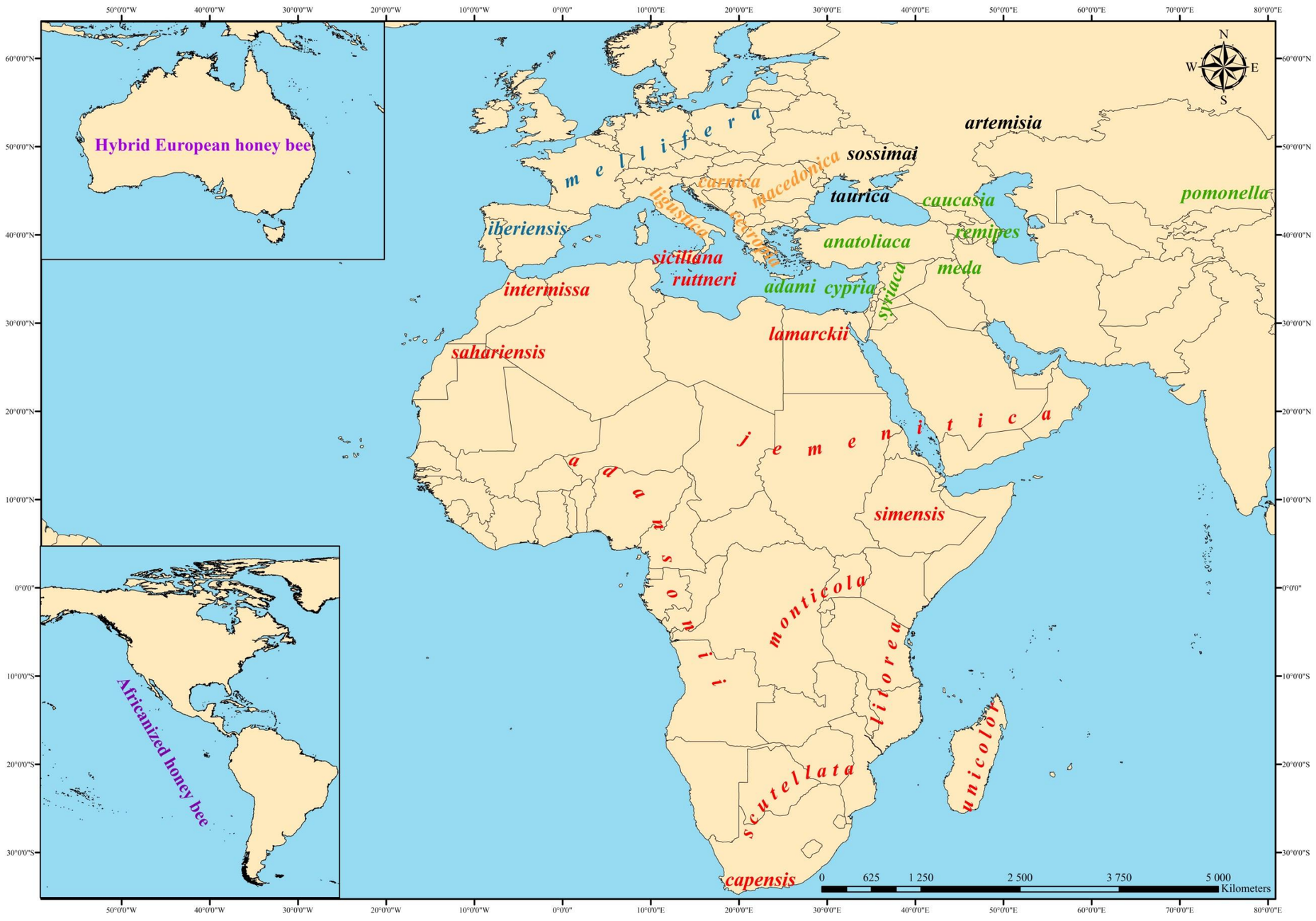


Figure II - 4. Geographical distribution of 30 *Apis mellifera* subspecies. The color of the names indicate the evolutionary lineage: A - African lineage (red), M - Western and Northern European lineage (blue), C – Eastern European lineage (orange), and O – Middle East and Western Asia lineage (green). *A. m. sossimai* (Engel 1999), *A. m. taurica* (Alpatov 1938) and *A. m. artemisia* (Engel 1999) are subspecies with sparse information on their evolutionary lineage.

- *Apis mellifera scutellata* Lepeletier (1836), the African honey bee, is a subspecies that has a large distribution in Africa and ranges from South Africa northward along the eastern half of the continent to about Ethiopia.
- *Apis mellifera capensis* Escholtz (1822), the Cape honey bee, exclusively occurs in the Cape Region of the South Africa.
- *Apis mellifera monticola* Smith (1961), the East African Mountain honey bee, is a subspecies that inhabits within the mountains of Eastern Africa (Kenya and Tanzania).
- *Apis mellifera unicolor* Latreille (1804), the Malagasy honey bee, is exclusively distributed in Madagascar.
- *Apis mellifera litorea* Smith (1961), the East African honey bee, is distributed along the eastern coast of tropical Africa occurring from Southern Kenya (perhaps even the southernmost portions of Somalia) to Mozambique.
- *Apis mellifera ruttneri* Sheppard *et al* (1997), the Maltese honey bee, is only distributed on the island of Malta in the Mediterranean sea.
- *Apis mellifera simensis* Meixner *et al*. (2011), the Ethiopian Mountain honey bee, is distributed in the mountain systems of Ethiopia.

Eastern Europe lineage (C)

- *Apis mellifera ligustica* Spinola (1806), the Italian honey bee or yellow bee, is distributed along the Italian Peninsula and confined by the Alps to the north and the Mediterranean Sea southwards.
- *Apis mellifera carnica* Pollmann (1879), the Carniolan honey bee, is a subspecies that occurs in Austria, Slovenia, Croatia, Bosnia-Herzegovina, Albania, Serbia, Hungary and Romania.
- *Apis mellifera macedonica* Ruttner (1988), the Macedonian honey bee, extends across Bulgaria, the Former Yugoslav Republic of Macedonia (FYROM), Greece, Romania, Ukraine and Turkey.
- *Apis mellifera cecropia* Kiesenwetter (1860), the Greek honey bee, is distributed all over Southern Greece, including the Peloponnese and surrounding Aegean islands.

Near East and Asia lineage (O)

- *Apis mellifera syriaca* Skorikov (1929), the Syrian honey bee or Palestine honey bee, is distributed along the eastern shores of the Mediterranean Sea; north from Syria to the Negev Desert in the south (Syria, Lebanon, Jordan and Israel).
- *Apis mellifera remipes* Gerstäcker (1862), the Yellow Armenian honey bee, is a subspecies only distributed in Armenia.
- *Apis mellifera meda* Skorikov (1929), the Median honey bee, is distributed in Iran and Iraq but does range into Southeastern Turkey and Northern Syria.
- *Apis mellifera anatoliaca* Maa (1953), the Anatolian honey bee, is a subspecies that occurs throughout Turkey.
- *Apis mellifera caucasia* Pollmann (1889), the Caucasian honey bee, is a subspecies that occurs in the Caucasus Mountains.
- *Apis mellifera cypria* Pollmann (1879), the Cyprian honey bee, is a subspecies distributed only in the island of Cyprus.
- *Apis mellifera adami* Ruttner (1975), the Cretan honey bee, is a subspecies that occurs on the island of Crete in the Mediterranean Sea.
- *Apis mellifera pomonella* Sheppard & Meixner (2003), the Tien Shan honey bee, is endemic of the Tien Shan Mountains in Central Asia along east-west orientation from south central Kazakhstan to western China.

The Iberian honey bee (*Apis mellifera iberiensis*)

The Iberian Peninsula provides one of the most interesting settings in Europe for studying contact zones. High geological, physiographical and climatic complexity and diversity, together with isolation from Europe and proximity to Africa (especially at the Strait of Gibraltar), made this southernmost European region an important refuge during the Quaternary glaciations (reviewed by Hewitt 2000) and a bridge, for the more vagile organisms, between the two continents (Carranza *et al*. 2004; Cosson *et al*. 2005; Guillaumet *et al*. 2006; Whitfield *et al*. 2006; Wallberg *et al*. 2014). These features made Iberia not only a place of divergence during periods of isolation but also a contact zone during periods of expansion, as reported for a wide array of plant and animal taxa (extensively reviewed by Weiss & Ferrand 2007), including the Iberian honey bee: *Apis mellifera iberiensis*.

The Iberian honey bee has been one of the most intensively surveyed honey bee subspecies in order to understand its complex pattern of diversity, typical of glacial refuges. Several

phylogeographical studies have been performed based on morphology and behavior (Ruttner 1988, Cornuet & Fresnaye 1989; Arias *et al*. 2006; Miguel *et al*. 2011), pheromones (Hepburn & Radloff 1996), allozymes (Nielsen *et al*. 1994; Smith & Glenn 1995; Arias *et al*. 2006), mitochondrial DNA (Smith *et al*. 1991; Garnery *et al*. 1995, 1998a; De la Rúa *et al*. 2004a, b, 2005; Miguel *et al*. 2007; Cánovas *et al*. 2008; Pinto *et al.* 2013), microsatellites (Franck *et al*. 1998; Garnery *et al*. 1998b; Miguel *et al*. 2007; Miguel *et al.* 2011; Cánovas *et al*. 2011) and SNPs (Whitfield *et al*. 2006; Han *et al*. 2012).

Differential and complex diversity patterns from the numerous biparental and maternal surveys of Iberian honey bees and the underlying processes shaping genetic variation remain controversial. Arguments based on selection, demography, and contemporary human-mediated processes have been favored by different authors. Early phylogeographical studies of morphology (Ruttner *et al*. 1978; Cornuet & Fresnaye 1989) and the allozyme malate dehydrogenase (Nielsen *et al*. 1994; Smith & Glenn 1995; Arias *et al*. 2006) revealed the existence of a smooth gradient extending from North Africa to France with Iberian honey bees showing intermediate phenotypes. This pattern raised the hypothesis of primary intergradation and an African origin for this subspecies (Ruttner *et al*. 1978). However, mitochondrial polymorphisms showed the co-occurrence of highly divergent A-derived and M-derived haplotypes forming not a smooth but a steep south–north cline in the Iberian Peninsula (Garnery *et al*. 1995; Franck *et al*. 1998; Arias *et al*. 2006; Miguel *et al*. 2007; Cánovas *et al*. 2008; Pinto *et al*. 2013), a pattern that was more consistent with a secondary contact scenario (Smith *et al*. 1991). Adding to the complexity, microsatellites (Franck *et al*. 1998; Garnery *et al*. 1998b; Cánovas et al. 2011; Miguel *et al*. 2011) and recent geometric morphometric data (Miguel *et al*. 2011) exhibited virtually no differentiation and no traces of African genes in Iberian honey bee populations and revealed a sharp break between Iberian and northern African populations, thereby supporting neither hypothesis. The difference between maternal and bi-parental variation led Franck *et al*. (1998) to reject the secondary contact hypothesis and propose historical human-assisted introductions of African colonies with selection the best explanation for the reported morphological and allozymic clines and the diffusion and maintenance of African haplotypes in the south-western half of the Iberian Peninsula. A recent study using SNPs suggests, however, that while selection may have shaped the genome of lineage-M honey bees, the process occurred during ancient expansions from Africa into Western Europe, resurrecting the primary intergradation hypothesis (Zayed & Whitfield 2008). All these arguments highlight to the Iberian honey bee as an interesting model organism for

evolutionary studies in order to understand the different processes of speciation and local adaptation.

References

Adams J, Rothman ED, Kerr WE, Paulino ZL (1977). Estimation of the number of sex alleles and queen matings from diploid male frequencies in a population of *Apis mellifera*. *Genetics,* **86,** 583-596.

Alburaki M, Moulin S, Legout H, Alburaki A, Garnery L (2011) Mitochondrial structure of Eastern honeybee populations from Syria, Lebanon and Iraq. *Apidologie*, **42**, 628–641.

Alpatov WW (1938) Contribution to the study of variation in the honeybee, VI. Carniolan and Crimean bees and their places among other forms of *Apis mellifera*. *Zoologischeskiy Zhurnal*, **17**, 473-481.

Arias MC, Rinderer TE, Sheppard WS (2006) Further characterization of honey bees from the Iberian Peninsula by allozyme, morphometric and mtDNA haplotype analyses. *Journal of Apicultural Research* **45**, 188-196.

Arias MC, Sheppard WS (2005) Phylogenetic relationships of honey bees (Hymenoptera:Apinae:Apini) inferred from nuclear and mitochondrial DNA sequence data. *Molecular Phylogenetics and Evolution*, **37**, 25–35.

Baldensperger PJ (1923) Sur l'abeille saharienne .6th congrès International d'Apiculture, Marseille 1922, 61-64.

Beye M, Gattermeier I, Hasselmann M *et al*. (2006) Exceptionally high levels of recombination across the honey bee genome. *Genome Research*, **16**, 1339–1344.

Beye M, Hasselmann M, Fondrk MK, Page RE, Omholt SW (2003). The gene *csd* is the primary signal for sexual development in the honeybee and encodes an SR-type protein. *Cell,* **114**, 419-429.

Bouga M, Harizanis PC, Kilias G, Alahiotis S (2005a) Genetic divergence and phylogenetic relationships of honey bee *Apis mellifera* (Hymenoptera: Apidae) populations from Greece and Cyprus using PCR - RFLP analysis of three mtDNA Segments. Apidologie, **36**, 335-344.

Bouga M, Kilias G, Harizanis PC, Papasotiropoulos V, Alahiotis S (2005b) Allozyme variability and phylogenetic relationships in honey bee (Hymenoptera: Apidae: *A. mellifera*) populations from Greece and Cyprus. Biochemical Genetics, **43**, 471-484.

Cánovas F, De la Rúa P, Serrano J, Galián J (2008) Geographical patterns of mitochondrial DNA variation in *Apis mellifera iberiensis* (Hymenoptera: Apidae). *Journal of Zoological Systematics and Evolutionary Research*, **46**, 24–30.

Cánovas F, De la Rúa P, Serrano J, Galián J (2011) Microsatellite variability reveals beekeeping influences on Iberian honeybee populations. *Apidologie*, **42**, 235–251.

Cardinal S, Danforth BN (2011) The antiquity and evolutionary history of social behavior in bees. PLoS ONE, 6, e21086.

Carranza S, Arnold EN, Wade E, Fahd S (2004) Phylogeography of the false smooth snakes, *Macroprotodon* (Serpentes, Colubridae): mitochondrial DNA sequences show European populations arrived recently from Northwest Africa. *Molecular Phylogenetics and Evolution*, **33**, 523–532.

Cockerell TDA (1906) New Rocky Moutain bees and other notes. *Canadian Entomologist*, **38**, 160-166.

Cornuet JM, Fresnaye J (1989) Biometrical study of honey bee populations from Spain and Portugal. *Apidologie,* **20**, 93-101.

Cornuet JM, Garnery L (1991) Mitochondrial DNA variability in honeybees and its phylogeographic implications. *Apidologie*, **22**, 627–642.

Cosson JF, Hutterer R, Libois R *et al*. (2005) Phylogeographical footprints of the Strait of Gibraltar and Quaternary climatic fluctuations in the western Mediterranean: a case study with the greater white-toothed shrew, *Crocidura russula* (Mammalia: Soricidae). *Molecular Ecology*, **14**, 1151–1162.

De la Rúa P, Hernandez-Garcia R, Jiménez Y, Galián J, Serrano J (2005) Biodiversity of *Apis mellifera iberica* (Hymenoptera: Apidae) from northeastern Spain assessed by mitochondrial analysis. *Insect Systematics & Evolution*, **36**, 21–28.

De la Rúa P, Hernandez-Garcia R, Pedersen BV, Galián J, Serrano J (2004a) Biodiversity of *Apis mellifera iberica* (Hymenoptera: Apidae) from northeastern Spain assessed by mitochondrial analysis. *Insect Systematics & Evolution*, **36**, 21–28.

De la Rúa P, Jaffé R, Dall'Olio R, Muñoz I, Serrano J (2009) Biodiversity, conservation and current threats to European honeybees. *Apidologie*, **40**, 263–284.

De la Rúa P, Jiménez Y, Galián J, Serrano J (2004b) Evaluation of the biodiversity of honey bee (*Apis mellifera*) populations from eastern Spain. *Journal of Apicultural Research*, **43**, 162–166.

Elsik CG, Mackey AJ, Reese JT *et al.* (2007) Creating a honey bee consensus gene set. *Genome Biology*, 8, R13.

Elsik CG, Worley KC, Bennett AK *et al.* (2014) Finding the missing honey bee genes: lessons learned from a genome upgrade. *BMC Genomics*, **15**, 86.

Engel MS (1999) The taxonomy of recent and fossil honey bees (Hymenoptera : Apidae; *Apis*). *Journal of Hymenoptera Research*, **8**, 165–196.

Franck P, Garnery L, Celebrano G, Solignac M, Cornuet JM (2000) Hybrid origins of honey bees from Italy (*Apis mellifera ligustica*) and sicily (*A. m. sicula*). *Molecular Ecology*, **9**, 907–921.

Franck P, Garnery L, Loiseau A *et al.* (2001) Genetic diversity of the honey bee in Africa: microsatellite and mitochondrial data. *Heredity*, **86**, 420–430.

Franck P, Garnery L, Solignac M, Cornuet JM (1998) The origin of west European subspecies of honeybees (*Apis mellifera*): new insights from microsatellite and mitochondrial data. *Evolution*, **52**, 1119–1134.

Gallai N, Salles JM, Settele J, Vaissière BE (2009) Economic valuation of the vulnerability of world agriculture confronted with decline polinator. *Ecological Economics*, **68**, 810-821.

Garnery L, Cornuet JM, Solignac M (1992) Evolutionary history of the honey bee *Apis mellifera* inferred from mitochondrial DNA analysis. *Molecular Ecology*, **1**, 145–154.

Garnery L, Franck P, Baudry E *et al.* (1998a) Genetic diversity of the west European honey bee (*Apis mellifera mellifera* and *A. m. iberica*). I. Mitochondrial DNA. *Genetics Selection Evolution* **30**, S31-S47.

Garnery L, Franck P, Baudry E *et al.* (1998b) Genetic diversity of the west European honey bee (*Apis mellifera mellifera* and *A. m. iberica*). II. Microsatellite loci. *Genetics Selection Evolution* **30**, S49-S74.

Garnery L, Mosshine EH, Oldroyd BP, Cornuet JM (1995) Mitochondrial DNA variation in Moroccan and Spanish honey bee populations. *Molecular Ecology*, **4**, 465–471.

Garnery L, Solignac M, Celebrano G, Cornuet JM (1993) A simple test using restricted PCR-amplified mitochondrial DNA to study the genetic structure of *Apis mellifera* L. *Experientia*, **49**, 1016–1021.

Gerstäcker CEA (1862) Über die geographische Verbreitung und die Abänderungen der Honigbiene nebst Bemerkungen über die auländischen Honigbienen der alten Welt. Festchrift XI Wander-Versammlung deutscher Bienenwirthe, Potsdam, Germany 75 pp.

Grassi B (1880) Saggio di una Monografia delle Api d'Italia, Milano [Milan], Italy.

Guillaumet A, Pons JM, Godelle B, Crochet PA (2006) History of the Crested Lark in the Mediterranean region as revealed by mtDNA sequences and morphology. *Molecular Phylogenetics and Evolution*, **39**, 645–656.

Han F, Wallberg A, Webster MT (2012) From where did the Western honeybee (*Apis mellifera*) originate? *Ecology and Evolution*, **2**, 1949–1957.

Hepburn HR, Radloff SE (1996) Morphometric and pheromonal analyses of *Apis mellifera* L along a transect from the Sahara to the Pyrenees. Apidologie, **27**, 35-45.

Hernández-Garcia R, De la Rúa P, Serrano J (2009) Mating frequency in *Apis mellifera iberiensis* queens. *Journal of Apicultural Research and Bee World*, **48**, 121-125.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. Nature, 405, 907–913.

Honeybee Genome Sequencing Consortium (2006) Insights into social insects from the genome of the honey bee *Apis mellifera*. *Nature*, **443**, 931-949.

Kandemir I, Kence M, Sheppard WS, Kence A (2006) Mitochondrial DNA variation in honey bee (*Apis mellifera* L.) populations from Turkey. *Journal of Apicultural Research*, **45**, 33–38.

Kandemir I, Özkan A, Fuchs S (2011) Reevaluation of honey bee (*Apis mellifera*) microtaxonomy: a geometric morphometric approach. *Apidologie*, 42, 618–627.

Kerr WE, Zucchi R, Nakaida JT, Butolo JE (1962). Reproduction in the social bees (Hymenoptera: Apidae). *Journal of the New York Entomological Society,* **70**, 265-276.

Kiesenwetter EAH von (1860) Ueber die Bienen des Hymettus. *Berliner Entomologische Zeitschrift*, **4**, 315-317.

Kotthoff U, Wappler T, Engel MS (2013) Greater past disparity and diversity hints at ancients migrations of European honey bee lineages into Africa and Asia. *Journal of Biogeography*, **40**, 1832-1838.

Latreille PA (1804) Mémoire sur un Gâteau de Ruche d'une Abeille des Grandes-Indes, et sur les Différences des Abeilles proprement dites, vivant en grande Société, de l'ancien Continent et du nouveau. *Annalies du Musée d'Histoire Naturelle*, **4**,383-394.

Le Conte Y, Navajas M (2008) Climate change: impact on honey bee populations and diseases. *Revue scientifique et technique (International Office of Epizootics)*, **27**, 499-510.

Lepeletier (1836) Histoire Naturelle des Insectes. Suites à Buffon. Hyménoptères, vol. 1. Roret, Paris, France, 547pp.

Linnaeus C (1758) Systema Naturae per Regna Tria Naturae, Secundum Classes, Ordines, Genera, Species, cum Characterisbus, Differentiis, Synonymis, Locis, ed 10, vol.1 Reformtat. Salviae, Holmiae [Stockholm], Sweden, 824 pp.

Maa TC (1953) An inquiry into systematics of the tribus Apidini or honeybees (Hym). *Treubia*, **21**, 525-640.

Meixner MD, Leta MA, Koeniger N, Fuchs S (2011) The honey bees of Ethiopia represent a new subspecies of *Apis mellifera-Apis mellifera simensis* n. ssp. *Apidologie*, **42**, 425–437.

Mexiner MD, Pinto MA, Bouga M *et al*. (2013) Standard methods for characterising subspecies and ecotypes of *Apis mellifera*. Journal of Apicultural Research, **52**, 1-27.

Miguel I, Baylac M, Iriondo M *et al.* (2011) Both geometric morphometric and microsatellite data consistently support the differentiation of the *Apis mellifera* M evolutionary branch. *Apidologie,* **42**, 150-161.

Miguel I, Iriondo M, Garnery L, Sheppard WS, Estonba A (2007) Gene flow within the M evolutionary lineage of *Apis mellifera*: role of the Pyrenees, isolation by distance and post-glacial re-colonization routes in the western Europe. *Apidologie,* **38**, 141-155.

Nielsen D, Page RE, Jr., Crosland MW (1994) Clinal variation and selection of MDH allozymes in honey bee populations. *Experientia,* **50**, 867-871.

Palmer MR, Smith DR, Kaftanoğlu O (2000) Turkish honey bees: genetic variation and evidence for a fourth lineage of *Apis mellifera* mtDNA. *Journal of Heredity,* **91**, 42-66.

Pinto MA, Henriques D, Neto M *et al*. (2013) Maternal diversity patterns of Ibero-Atlantic populations reveal further complexity of Iberian honey bees. *Apidologie*, **44**, 430–439.

Pollmann A (1879) Werth der verschiedenen Bienenracen und deren Varietäten, bestimmt durch Urtheile namhafter Bienenzünchter, Voight, Leipzig, Germany, 69pp.

Pollmann A (1889) Werth der verschiedenen Bienenracen und deren Varietäten, bestimmt durch Urtheile namhafter Bienenzünchter, Voight, Leipzig, Germany, viii +100 pp.

Raffiudin R, Crozier RH (2007) Phylogenetic analysis of honey bee behavioral evolution. *Molecular Phylogenetics and Evolution*, **43**, 543–552.

Roberts WC (1944). Multiple mating of queen bees proved by progeny and flight tests. *Gleanings in Bee Culture,* **72**, 255-260.

Ruttner F (1975) Die kretische Biene, *Apis mellifera adami*. Allgemeine deutche Imkerzeitung, **9**, 271-272.

Ruttner F (1988) Biogeography and Taxonomy of Honey Bees. Springer, Berlin.

Ruttner F, Tassencourt L, Louveaux J (1978) Biometrical-statistical analysis of the geographic variability of *Apis mellifera* L. *Apidologie,* **9**, 363-381.

Escholtz JF (1822) Entomographien, vol.1. Reimer. Berlin, Germany, 59-186 pp.

Shaibi T, Muñoz I, Dall'Olio R *et al.* (2009) *Apis mellifera* evolutionary lineages in Northern Africa: Libya, where orient meets occident. *Insectes Sociaux*, **56**, 293-300.

Sheppard WS, Arias MC, Grech A, Meixner MD (1997) *Apis mellifera ruttneri*, a new honey bee subspecies from Malta. *Apidologie,* **28**, 287-293.

Sheppard WS, Meixner MD (2003) *Apis mellifera pomonella*, a new honey bee subspecies from Central Asia. *Apidologie,* **34**, 367-375.

Sinacori A, Rinderer TE, Lancaster V, Sheppard WS (1998) A morphological and mitochondrial assessment of *Apis mellifera* from Palermo, Italy. *Apidologie*, **29**, 481–490.

Skorikov AS (1929) Eine neue Basis für eine Revision der Gattung *Apis* L. *Reports on Applied Entomology*, **4**, 1-59.

Smith F (1861) Description of new speces of hymenopterous insects collected by Mr A. R. Wallace at Celebes. *Proceedings of the Linnean Society of London*, **5**, 57-93.

Smith DR, Glenn TC (1995) Allozyme polymorphisms in spanish honeybees (*Apis mellifera iberica*). *Journal of Heredity,* **86**, 12-16.

Smith DR, Palopoli MF, Taylor BR *et al.* (1991) Geographical overlap of two mitochondrial genomes in Spanish honeybees (*Apis mellifera iberica*). *Journal of Heredity,* **82**, 96-100.

Smith DR, Slaymaker A, Palmer M, Kaftanoğlu O (1997) Turkish honey bees belong to the east Mediterranean mitochondrial lineage. *Apidologie*, **28**, 269-274.

Spinola M (1806) Insectorum Liguriai Species Novae aut Rariores, Quas in Agro Ligustico nuper Detexit, Descripsit, et Iconibus Illustravit Maximilianus Spinola, adjecto Catologo Specirum Auctoribus jam Enumeratum, Quae un Aedem Regione Passim Ocuurrunt, vol. 1 Gravier, Genuae [Genoa], Italy, xvii+159+[1] pp., 2 pls.

Tarpy DR, Page RE (2000). No behavioral control over mating frequency in queen honey bees (*Apis mellifera* L.): Implications for the evolution of extreme polyandry. *The American Naturalist,* **155**, 820-827.

Wallberg A, Han F, Wellhagen G *et al.* (2014) A worldwide survey of genome sequence variation provides insight into the evolutionary history of the honeybee *Apis mellifera*. *Nature Genetics*, **46**, 1081–1088.

Weiss S, Ferrand N (2007) Current perspectives in phylogeography and the significance of South European refugia in the creation and maintenance of European biodiversity. In: Phylogeographyof Southern European Refugia (eds Weiss S, Ferrand N), pp. 237–255. Springer, The Netherlands.

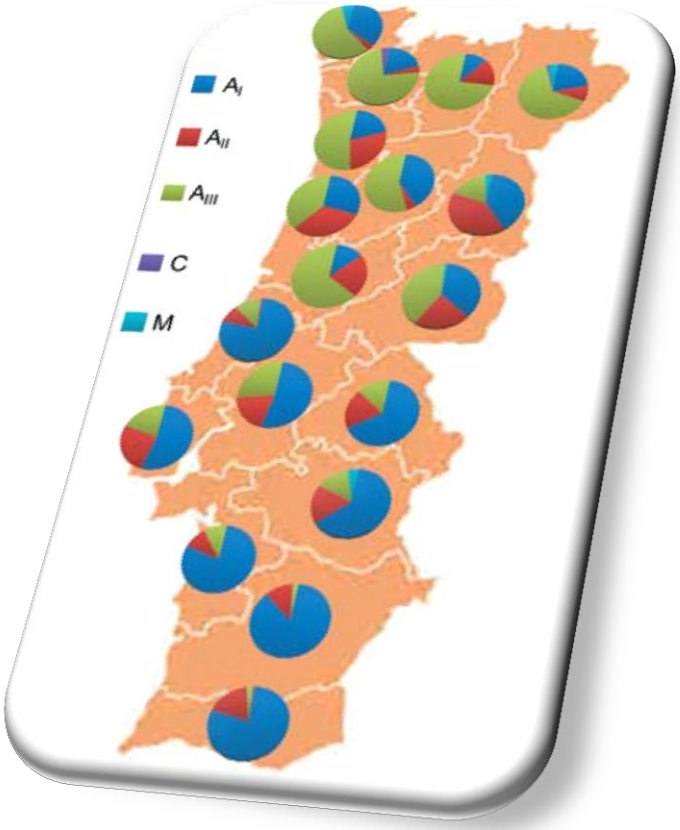
Whitfield CW, Behura SK, Berlocher SH *et al*. (2006) Thrice out of Africa: ancient and recent expansions of the honey bee, *Apis mellifera*. *Science*, **314**, 642–645.

Wilson EO (1971) The insect societies. Belknap Press, Harvard.

Winston M (1987) The biology of the Honey Bee. Harvard University Press, London.

Woyke J (1963). What happens to diploid drone larvae in a honeybee colony. *Journal of. Apicultural Research,* **2**, 73-75.

Zayed A, Whitfield CW (2008) A genome-wide signature of positive selection in ancient and recent invasive expansions of the honey bee *Apis mellifera*. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 3421-3426.



Chapter III

The Atlantic side of the Iberian Peninsula: a hot-spot of novel African honey bee maternal diversity

The work presented in this chapter has been published:

Pinto MA, Muñoz I, **Chávez-Galarza J**, De la Rúa P (2012) The Atlantic side of the Iberian Peninsula: a hot-spot of novel African honey bee maternal diversity. **Apidologie**, 43: 663-673

Abstract

The Iberian honey bee has been one of the most surveyed subspecies for genetic diversity. Yet, previous studies have missed an important component of Iberian honey bee variation harbored by populations inhabiting the Atlantic side of the Iberian Peninsula. Herein, we provide a fuller picture of the Iberian honey bee maternal diversity by revealing 16 novel haplotypes detected in honey bees from Portugal. Of the 16 haplotypes, all of African ancestry, 15 belong to the Atlantic sub-lineage A_{III} while only one fits the most common sub-lineage A_{I}. This level of new variation is remarkable as it represents a 59% increase in the wide-range African lineage and a 188% in sub-lineage A_{III}. Our findings further highlight the complexity of the Iberian honey bee diversity patterns and reinforce the importance of this southernmost European territory as a reservoir of *Apis mellifera* genetic diversity, a resource increasingly important in a rapidly changing and demanding world.

Keywords: Iberian honey bee, genetic diversity, mtDNA, *Dra*I test, Portugal

Introduction

Understanding patterns and underlying processes of diversity of the western honey bee (*Apis mellifera* L.) has been a major goal of numerous genetic studies, which in a population declining scenario is becoming increasingly important because it can provide a stronger scientific basis for management and conservation decisions. Among the 29 recognized honey bee subspecies (Engel 1999; Sheppard & Meixner 2003), the Iberian honey bee has been one of the most intensively surveyed for diversity patterns. Indeed, studies of Iberian honey bees have been performed since the 1970s (Ruttner *et al*. 1978) using morphology (Cornuet & Fresnaye 1989; Arias *et al*. 2006; Miguel *et al*. 2011), allozymes (Smith & Glenn 1995; Arias *et al*. 2006), mitochondrial DNA (Smith *et al*. 1991; Garnery *et al*. 1995, 1998a; Franck *et al*. 1998; De la Rúa *et al*. 2001, 2004, 2005; Arias *et al*. 2006; Miguel *et al*. 2007; Cánovas *et al*. 2008), and microsatellites (Franck *et al*. 1998; Garnery *et al*. 1998b; De la Rúa *et al*. 2002, 2003; Miguel *et al*. 2007; Cánovas *et al*. 2011). Differential and complex patterns of diversity have emerged from these surveys, which have yet to be fully resolved.

The maternally inherited mitochondrial DNA (mtDNA) has been the marker of choice for assessing Iberian honey bee variation, particularly the PCR–RFLP of the intergenic $tRNA^{leu}$-cox2 region, also known as the *Dra*I test (Garnery *et al*. 1993). Over 2,500 colonies, mostly sampled in the eastern half of the Iberian Peninsula, have been screened with the *Dra*I test (Garnery *et al*. 1993, 1995, 1998a; Miguel *et al*. 2007; Cánovas *et al.* 2008). The data generated by this massive sampling confirmed early findings of coexistence of African (A) and western European (M) lineages, forming a south–north cline (Smith *et al*. 1991), and revealed unparalleled levels of haplotype diversity and complexity (Franck *et al*. 1998; Cánovas *et al*. 2008). Accordingly, it has been suggested that, as for other animal and plant species, the Iberian Peninsula served as a glacial refuge (Hewitt 1999, 2001; Arias *et al*. 2006; Gómez & Lunt 2007), and as a place of secondary contact between European and African evolutionary lineages (Smith *et al*. 1991; Garnery *et al.* 1995; Cánovas *et al*. 2008), and therefore it has been a stage for historical processes possibly involving recurring phenomena of local adaptation, contraction, fragmentation, expansion, and admixture.

In contrast with the populations inhabiting the eastern side of the Iberian Peninsula, Portuguese honey bees have been largely under-sampled. Yet, the few samples collected in Portugal (Garnery *et al*. 1998a; Arias *et al*. 2006; Miguel *et al*. 2007) suggest that the Atlantic side of Iberia may harbor an important component of the Iberian honey bee maternal diversity.

Therefore, a fuller understanding of diversity patterns of the Iberian honey bee requires further surveys of Ibero-Atlantic populations. As part of an ongoing genetic study of the honey bees occupying the Portuguese territory, we have detected 16 novel haplotypes with the *Dra*I test. This finding is non-trivial as this small corner of southern Europe is contributing with 59% of new variation to the wide-range African lineage. Herein, the novel haplotypes are fully described by the RFLP approach and by sequence data. Our findings suggest that the Atlantic side of the Iberian Peninsula harbors important genetic resources, especially in face of the escalating threats to the honey bee diversity.

Methods

Samples and RFLP analysis

As part of an ongoing study of the Portuguese honey bee populations, over 950 stationary colonies were sampled, between 2008 and 2010, covering every district of continental Portugal and the archipelagos of Azores and Madeira (sample sizes and locations are provided in Table Sup III - 1). Honey bee workers were collected from the inner part of the hives, placed in absolute ethanol, and then stored at −20°C until molecular analysis. The maternal ancestry of the 950 workers, each representing a single colony and a single apiary, was assessed using the *Dra*I test (Garnery *et al*. 1993), which consists on PCR amplification of the $tRNA^{leu}$-cox2 intergenic region, using the primers E2 and H2 (see PCR details in Garnery *et al*. 1993), followed by digestion with the restriction enzyme *Dra*I (see digestion and gel electrophoresis details in Cánovas *et al*. 2008). Of the 950 individuals scored using the complete set of restriction maps and restriction fragment sizes reported to date (De la Rúa *et al*. 1998, 2005; Franck *et al*. 2001; Collet *et al*. 2006), 43 exhibited a total of 16 unreported PCR–RFLP patterns. Location of the 43 colonies, from which the individuals were sampled, is shown in Fig. III - 1. In this study, the 16 novel PCR–RFLP patterns were fully characterized by sequencing.

Sequencing and sequence analysis

The $tRNA^{leu}$-cox2 intergenic region analyzed herein contains a non-coding sequence which size depends on the forms of the P element and the number of repeats of the Q element. In the African lineage, the P element can display two different forms: P_0 (sub-lineage A_I and A_{II}) and P_1 (sub-lineage A_{III}). The P_0 differs from P_1 by a 15-bp deletion. The Q element can be repeated in tandem

one to four times. A further distinction between sub-lineages A_I and A_{II} is provided by the number of *Dra*I recognition sites in the region spanning $tRNA^{leu}$ and 5′ end of the first Q element, with two and one sites, respectively (Garnery *et al*. 1993; De la Rúa *et al*. 1998; Franck *et al*. 1998).

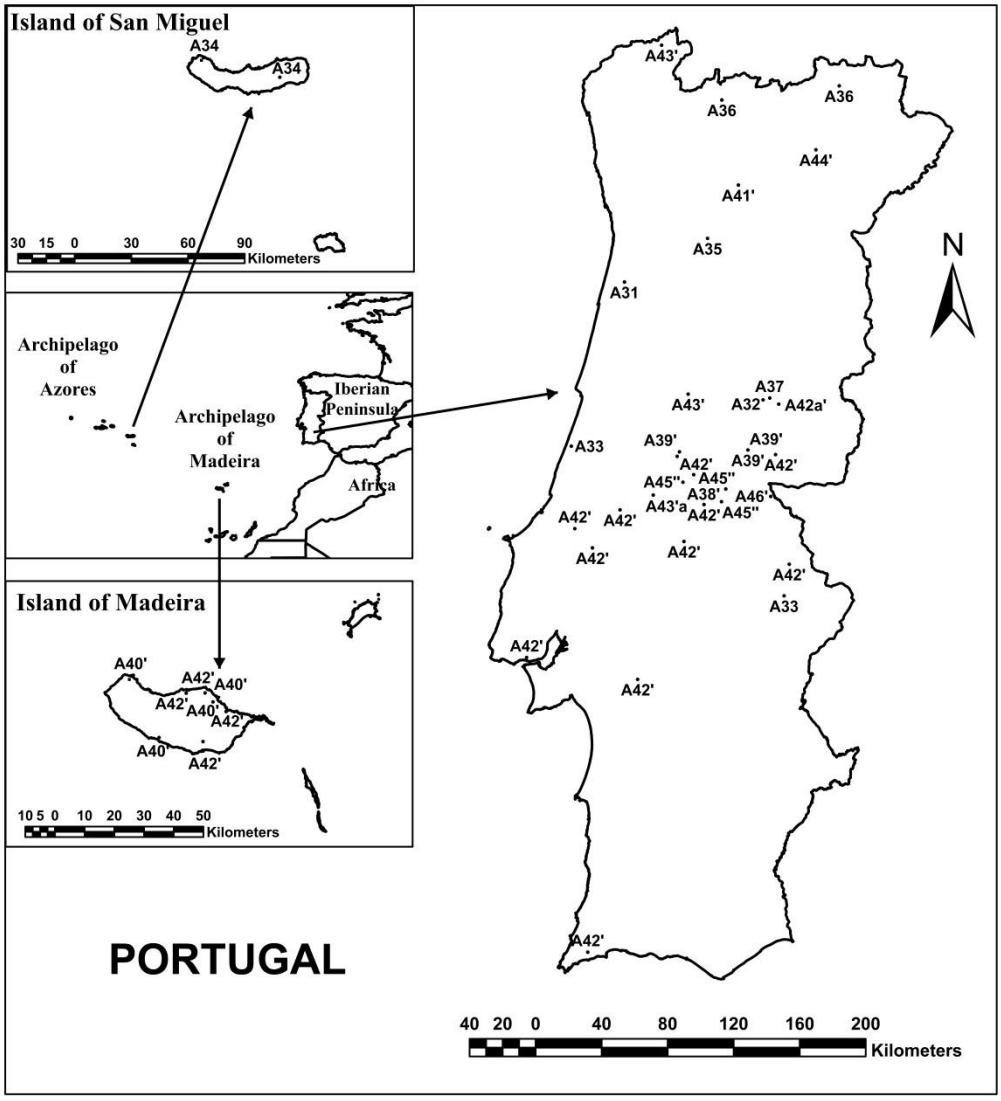


Figure III - 1. Location of the 43 colonies harboring the 16 novel PCR-RFLP haplotypes identified with the *Dra*I test.

The 16 novel PCR–RFLP patterns were further examined by sequencing the $tRNA^{leu}$-cox2 intergenic region for 20 individuals. In addition, three individuals exhibiting the previously reported band patterns, A3 (sub-lineage A_I), A14, and A16 (sub-lineage A_{III}), were also sequenced and included for comparison. Following PCR amplification, PCR products were purified either with isopropanol and ammonium acetate or using a column-based purification kit (Zymo Research®) and sent to Secugen S. L. (Madrid, Spain) or Macrogen (Seoul, Korea) for direct sequencing in both directions with primers E2 and H2. The sequences were checked for base calling using Seq-Man®

version 7.0.0 and then deposited in GenBank (http://www.ncbi.nlm.nih.gov). The 23 sequences, plus the published sequences of haplotypes A29a (FJ890930.1; Szalanski & Magnus 2010) and A30 (EF033654.1; Collet *et al*. 2006), belonging to sub-lineage A_{III}, were aligned using MEGA version 5.03 (Tamura *et al*. 2011). Proximity among the haplotypes was established using the median-joining network algorithm (Bandelt *et al*. 1999) as implemented in the program Network version 4.6.0.0 (Fluxus Engineering, Clare, UK; http://www.fluxus-engineering.com), with epsilon set to zero and downweighting the four most variable sites (weight=0 for the most variable site; weight=5 for the remaining three most variable sites). Indels (e.g., P_1 element, third and fourth Q element, 15-deletion of A40′) were considered as a single mutational step, being therefore coded as a 1-bp gap. Variable characters within the first (Q1), second (Q2), and third Q (Q3) elements were included in the analysis.

Results

The *Dra*I test performed on colonies surveyed across Portugal (continent and archipelagos of Azores and Madeira) identified 16 novel PCR–RFLP patterns (haplotypes) carried by 43 colonies (Fig. III - 1). The restriction maps and length of restriction fragments (Fig. III - 2) suggest that they all fit within the African evolutionary lineage (A). Following the nomenclature established earlier (Garnery *et al*. 1998a) and recently reviewed for lineage M (Rortais *et al*. 2011), the 16 haplotypes were numbered sequentially from A31 to A46' (accession numbers JQ746684–JQ746701). Haplotypes showing the same RFLP pattern but bearing three or four Q elements were further distinguished by addition of the symbols ' and '', respectively, after the haplotype number (Garnery *et al*. 1998a; Rortais *et al*. 2011).

Fig. III - 2 shows that the 16 patterns are dramatically distinct from those previously reported for the African lineage. The number of restriction fragments varied between three (A31, A41', A45'') and five (A32, A35, A37, A40'), with most haplotypes exhibiting a four-fragment pattern (A33, A34, A36, A38', A39', A42', A43', A44', A46'). The shortest band (28 bp) was displayed by four haplotypes (A35, A36, A37, A39') whereas the longest (1,064 bp) was unique to A45''.

Location of the 43 colonies, and corresponding distribution of the 16 haplotypes depicted in Fig. III - 1, shows that haplotype A42' was the most widespread and common (16 colonies) followed by A40' (four colonies). Haplotypes A39', A43', and A45'' were carried by three colonies whereas A33, A34, and A36 by two colonies. The remaining eight haplotypes were singletons (A31,

A32, A35, A37, A38', A41', A44', and A46'). While most haplotypes were detected in the center of continental Portugal, A34 and A40' were private to the islands of San Miguel (Azores) and Madeira, respectively. A42' was the only haplotype detected in both mainland and island (Madeira) populations.

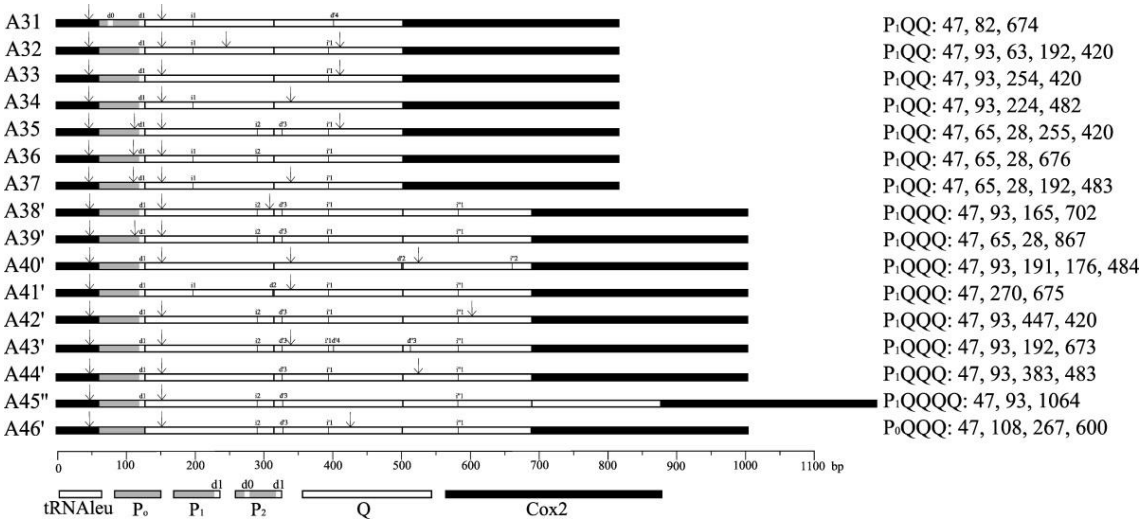


Figure III - 2. Restriction maps (left) and length of restriction fragments (right), deduced from *Dra*I restriction patterns and sequences of the tRNA^{leu}-cox2 intergenic region, of the 16 novel haplotypes found in Portugal. The *Dra*I recognition site is denoted with an arrow. Deletions and insertions are numbered preceded by the letters d and i, respectively. Deletion marked as d1 is specific of the P_1 element which defines sub-lineage A_{III}, as in Franck *et al.* (2001).

Sequence data (Fig. III - 3) confirmed the novelty of the 16 haplotypes of African ancestry and allowed identification of two additional variants of haplotypes A42' and A43', which were distinguished by a lowercase letter (A42'a, A43'a), as suggested by Rortais *et al.* (2011). Among the 16 haplotypes, 15 contained the P_1 element whereas only one (A46') exhibited the P_0 element. The length of the sequenced region ranged from 803 bp to 1,204 bp, depending on the number of Q elements and presence of indels. Fifteen haplotypes contained either two (A31, A32, A33, A34, A35, A36, A37) or three Q elements (A38', A39', A40', A41', A42', A43', A44', A46'). Only one haplotype (A45'') displayed a sequence with four Q elements (Fig. III - 2). As typical for this intergenic region, numerous indels of variable sizes were detected within the P and Q elements (Fig. III - 3). In addition to the 15-bp deletion characteristic of the P_1 element (marked as "d1" in Fig. III - 2 and 3), large indels were displayed by A31, A34, A40', and A41'. The largest fragment (33 bp, marked as "c" in Fig. III - 3) was inserted in the Q1 element of haplotype A34. This 33-bp insertion produced the longest Q element (227 bp) ever reported. The 15-bp deletion motif of A40' (marked as "l" in Fig. III - 3) and A41' (marked as "h" in Fig. III - 3) was a perfect match of d1, the

only difference being the position. Indeed, while d1 was located at the 3' end of the P element, deletions of A41' and A40' were at the 3' end of Q1 and Q2, respectively. In addition to the variation originated from large indels, there were 10 short indels (1–2 bp) and 17 single-base substitutions, of which nine resulted in a gain/loss of the *Dra*I recognition site. The most striking mutation, displayed by seven haplotypes, is the "AG" inserted at the 3' end of Q1 element (marked as "f" in Fig. III - 3). Interestingly, except for the Q3 of haplotype A41', no other Q element carried this mutation.

A median-joining network (Fig. III - 4) based on 36 variable sites (18 coded indels and 18 substitutions) illustrates the relationships among the novel and previously described haplotypes of sub-lineages A_I (A3) and A_{III} (A14, A16, A29a, A30). Two distinct clusters, separated by the number of Q elements, are represented in the network. The more poorly resolved right-hand side cluster connects the haplotypes with two Q elements. One unresolved connection and four hypothetical haplotypes (unsampled or extinct) link this group to A35, the closest haplotype of the left-hand side cluster. Haplotypes A36 and A37 are the closest in the group with only one mutational step (a transition in A36 that led to loss of *Dra*I restriction site) separating them. The haplotype pairs A30/A37 and A32/A33 are separated by one hypothetical haplotype and two mutational steps (one is a transversion that led to a gain/loss of *Dra*I restriction site). The previously described A14 and the novel A34, which are derived from the same hypothetical haplotype, are the most distant in the group. A minimum of eight mutations, including the duplication of the Q element, and two hypothetical nodes separate the closest haplotypes of both clusters (A35 and A43'a).

The better resolved left-hand side cluster connects the haplotypes with three and four Q elements. The 14 haplotypes are separated from each other by one to eight mutational steps and three hypothetical nodes. The previously described A29a is central to this group, showing considerably more connections (eight) than any other haplotype in the network. This central haplotype is connected to the closest ones by a single mutation (1-bp deletion in Q2 for A16, and a substitution in Q3 for A42'a) and to the most distant (A41') by eight mutational steps and one hypothetical node. The previously described A3 haplotype of sub-lineage A_I (as defined by Franck *et al.* 2001) and the novel A46' are together in a branch that is separated from A29a by three and four mutations, respectively. These two sub-lineage A_I haplotypes differ from each other by one transition mutation (C/T), which accounted for an additional *Dra*I recognition site in Q2 of haplotype A46' (substitution 12 in Fig. III - 3).

	76	77	78	79	80	81	82	83	84	85	86	102	111	112	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	158	197	198	199	200	201
	a											1	2	3	d1															4	c				
A3	A	T	A	A	A	A	T	A	A	A	A	T	T	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	A	-	-	-	-	-
A14	A	T	A	A	A	A	T	A	A	A	A	C	T	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-
A16	A	T	A	A	A	A	T	A	A	A	A	T	T	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-
A29a	A	T	A	A	A	A	T	A	A	A	A	T	T	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-
A30	A	T	A	A	A	A	T	A	A	A	A	T	T	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-
A31	-	-	-	-	-	-	-	-	-	-	-	T	T	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-
A32	A	T	A	A	A	A	T	A	A	A	A	T	T	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-
A33	A	T	A	A	A	A	T	A	A	A	A	T	T	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-
A34	A	T	A	A	A	A	T	A	A	A	A	C	A	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	T	T	T	A	T
A35	A	T	A	A	A	A	T	A	A	A	A	T	T	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-
A36	A	T	A	A	A	A	T	A	A	A	A	T	T	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-
A37	A	T	A	A	A	A	T	A	A	A	A	T	T	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-
A38'	A	T	A	A	A	A	T	A	A	A	A	T	T	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-
A39'	A	T	A	A	A	A	T	A	A	A	A	T	T	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-
A40'	A	T	A	A	A	A	T	A	A	A	A	T	T	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-
A41	A	T	A	A	A	A	T	A	A	A	A	T	T	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-
A42'	A	T	A	A	A	A	T	A	A	A	A	T	T	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-
A42'a	A	T	A	A	A	A	T	A	A	A	A	T	T	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-
A43'	A	T	A	A	A	A	T	A	A	A	A	T	T	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-
A43'a	A	T	A	A	A	A	T	A	A	A	A	T	T	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-
A44'	A	T	A	A	A	A	T	A	A	A	A	T	T	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-
A45''	A	T	A	A	A	A	T	A	A	A	A	T	T	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-
A46'	A	T	A	A	A	A	T	A	A	A	A	T	T	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	A	-	-	-	-	-

	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	252	275	303	304	326	327
																													d	5	6	e		7	f
A3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	A	-	-	G	A
A14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	C	A	T	G	-
A16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	C	-	-	G	A
A29a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	C	-	-	G	A
A30	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	T	C	-	-	G	-
A31	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	T	C	-	-	G	-
A32	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	A	C	-	-	G	-
A33	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	C	-	-	G	-
A34	T	T	T	T	A	T	A	T	T	T	A	A	T	A	A	A	A	T	A	A	A	T	A	A	T	T	A	A	-	T	C	-	-	G	-
A35	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	C	-	-	G	A
A36	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	T	C	-	-	G	-
A37	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	T	C	-	-	G	-
A38'	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	C	-	-	G	A
A39'	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	C	-	-	G	A
A40'	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	C	-	-	G	-
A41	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	T	C	-	-	A	-
A42'	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	C	-	-	G	A
A42'a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	C	-	-	G	A
A43'	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	C	-	-	G	A
A43'a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	C	-	-	G	A
A44'	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	C	-	-	G	A
A45''	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	C	-	-	G	A
A46'	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	A	-	-	G	A

Figure III - 3. Variable sites of tRNA^{leu}-cox2 intergenic region of the 16 novel haplotypes, the two additional sequence variants (A42'a and A43'a), and the previously reported African haplotypes (A3, A14, A16, A29a, A30). Sequences of A29a (FJ890930.1) and A30 (EF033654.1) were obtained from GenBank. Sequences of A3, A14, and A16 were obtained from individuals collected in Portugal. The numbers on the top line represent the nucleotide position within the tRNA^{leu}-cox2 sequenced region. Position 1 marks the beginning of the tRNA^{leu} gene, corresponding to position 3370 of the honey bee mitochondrial genome (Crozier & Crozier 1993). Substitution sites are numbered from 1 to 17. Indels (marked with a dash) are denoted by letters from a to q. The 15-bp deletion characteristic of the P$_1$ element is marked as d1 (position 115). Variable sites between positions 158–358, 365–550, and 551–714 are within Q1, Q2, and Q3 elements, respectively. Position 745 identifies the beginning of Q4. Full sequences are available in GenBank under accession numbers JQ746684–JQ746701.

	328	329	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	365	387	406	429	435	448	457	471	536	537	538	539	540	541	542	543	544
		g	h												8			i	9	10	j	k	11	12	13	l									
A3	G	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	-	G	C	T	T	T	C	C	A	T	T	T	A	A	T	T	T
A14	-	-	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	C	A	T	-	T	T	C	C	A	T	T	T	A	A	T	T	T
A16	G	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	-	G	T	-	T	T	C	C	A	T	T	T	A	A	T	T	T
A29a	G	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	-	G	T	T	T	T	C	C	A	T	T	T	A	A	T	T	T
A30	-	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	C	G	T	T	T	T	C	C	A	T	T	T	A	A	T	T	T
A31	-	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	C	G	T	T	-	T	C	C	A	T	T	T	A	A	T	T	T
A32	-	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	C	G	T	T	T	A	C	C	A	T	T	T	A	A	T	T	T
A33	-	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	C	G	T	T	T	A	C	C	A	T	T	T	A	A	T	T	T
A34	-	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	C	A	T	-	T	T	C	C	A	T	T	T	A	A	T	T	T
A35	G	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	-	G	T	T	T	A	C	C	A	T	T	T	A	A	T	T	T
A36	G	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	C	G	T	T	T	T	C	C	A	T	T	T	A	A	T	T	T
A37	-	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	C	A	T	T	T	T	C	C	A	T	T	T	A	A	T	T	T
A38'	G	A	A	T	T	T	A	A	T	T	T	A	T	T	T	A	A	-	G	T	T	T	T	C	C	A	T	T	T	A	A	T	T	T
A39'	G	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	-	G	T	T	T	T	C	C	A	T	T	T	A	A	T	T	T
A40'	-	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	C	A	T	-	T	T	C	C	-	-	-	-	-	-	-	-	-
A41	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	A	T	T	T	T	C	C	A	T	T	T	A	A	T	T	T
A42'	G	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	-	G	T	T	T	T	C	T	A	T	T	T	A	A	T	T	T
A42'a	G	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	-	G	T	T	T	T	C	C	A	T	T	T	A	A	T	T	T
A43'	G	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	-	A	T	T	-	T	C	C	A	T	T	T	A	A	T	T	T
A43'a	G	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	-	A	T	T	T	T	C	C	A	T	T	T	A	A	T	T	T
A44'	-	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	-	G	T	T	T	T	C	C	A	T	T	T	A	A	T	T	T
A45''	G	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	-	G	T	-	T	T	C	C	A	T	T	T	A	A	T	T	T
A46'	G	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	-	G	C	T	T	T	T	C	A	T	T	T	A	A	T	T	T

	545	546	547	548	549	550	551	557	579	593	621	640	709	713	714	745
							m	n	14	15	o	16	17	p		q
A3	A	T	T	A	A	A	A	C	G	C	T	T	A	-	-	-
A14	A	T	T	A	A	A	-	-	-	-	-	-	-	-	-	-
A16	A	T	T	A	A	A	A	C	G	C	T	T	A	-	-	-
A29a	A	T	T	A	A	A	A	C	G	C	T	T	A	-	-	-
A30	A	T	T	A	A	A	-	-	-	-	-	-	-	-	-	-
A31	A	T	T	A	A	A	-	-	-	-	-	-	-	-	-	-
A32	A	T	T	A	A	A	-	-	-	-	-	-	-	-	-	-
A33	A	T	T	A	A	A	-	-	-	-	-	-	-	-	-	-
A34	A	T	T	A	A	A	-	-	-	-	-	-	-	-	-	-
A35	A	T	T	A	A	A	-	-	-	-	-	-	-	-	-	-
A36	A	T	T	A	A	A	-	-	-	-	-	-	-	-	-	-
A37	A	T	T	A	A	A	-	-	-	-	-	-	-	-	-	-
A38'	A	T	T	A	A	A	A	C	G	C	T	T	A	-	-	-
A39'	A	T	T	A	A	A	A	C	G	T	T	T	A	-	-	-
A40'	-	-	-	-	-	-	A	C	A	C	-	T	A	A	G	-
A41	A	T	T	A	A	A	A	C	G	C	T	T	A	-	-	-
A42'	A	T	T	A	A	A	A	C	G	C	T	A	A	-	-	-
A42'a	A	T	T	A	A	A	A	C	G	C	T	A	A	-	-	-
A43'	A	T	T	A	A	A	A	-	G	C	T	T	A	-	-	-
A43'a	A	T	T	A	A	A	A	-	G	C	T	T	G	-	-	-
A44'	A	T	T	A	A	A	A	C	A	C	T	T	A	-	-	-
A45''	A	T	T	A	A	A	A	C	G	C	T	T	A	-	-	A
A46'	A	T	T	A	A	A	A	C	G	C	T	T	A	-	-	-

Figure III-3. (continued)

Discussion

This study further highlights the Atlantic side of the Iberian Peninsula as an important repository of Iberian honey bee maternal diversity. Analysis of Portuguese honey bee populations with the *Dra*I test revealed 16 novel haplotypes of African ancestry, which were confirmed by sequence data of the tRNA^{leu}-cox2 intergenic region. These haplotypes join the 27 African haplotypes previously described (De la Rúa *et al*. 1998, 2005; Franck *et al*. 2001; Collet *et al*. 2006) of which 17 have

been found in the Iberian Peninsula (Garnery *et al*. 1998a; De la Rúa *et al*. 2004, 2005; Cánovas *et al*. 2008), representing an increase of 59% and 94%, respectively. Most haplotypes (15) contained the P_1 element, typical of sub-lineage A_{III}, whereas only one was assigned to sub-lineage A_I, as defined by Franck *et al*. (2001). The 15 haplotypes were added to the eight previously reported for sub-lineage A_{III} (De la Rúa *et al*. 1998; Franck *et al*. 2001; Collet *et al*. 2006), representing an increase of 188%.

The Iberian honey bee has been one of the most intensively surveyed subspecies in its natural range (Garnery *et al*. 1995, 1998a; Franck *et al*. 1998; Arias *et al*. 2006; Miguel *et al*. 2007; Cánovas *et al*. 2008). Therefore, detection of such a remarkable number of novel haplotypes, mostly of sub-lineage A_{III} ancestry, was unexpected and suggests that prior studies have missed an important diversity component hold by the western populations of Iberia. This study not only adds to the complexity of the Iberian honey bee diversity patterns (Franck *et al*. 1998; Cánovas *et al*. 2008) but also reinforces the Atlantic distribution proposed for sub-lineage A_{III} (De la Rúa *et al*. 1998, 2006; Franck *et al*. 2001), as most colonies were detected in the north of continental Portugal (Fig. III - 1), which exhibits a more Atlantic climate contrasting with the more Mediterranean southern Portugal. Our findings further support an ancient natural colonization of the Iberian Peninsula by African swarms (De la Rúa *et al*. 2002, 2004; Cánovas *et al*. 2008) as the hypothesis of historical human-mediated multiple introductions (Franck *et al*. 1998; Garnery *et al*. 1998a) is untenable with such complex levels of diversity.

Among the 16 novel haplotypes, A42' is probably the oldest because of its higher frequency and wider geographical distribution (sole haplotype shared between mainland and island colonies). Alternatively, it could have been disseminated by human-assisted colony transportation, as occurred with A29. While the history of introductions of A29 (Collet *et al*. 2006; Prada *et al*. 2009) and its variant A29a (Szalanski & Magnus 2010) is unknown, we postulate that these haplotypes descend from colonies of Portuguese origin. The PCR–RFLP patterns of A29 (47/93/866) and A29a (47/93/867) are virtually indistinguishable from those of A16 (47/93/866). Additionally, sequence data shows that they are closely related (Fig. III - 4 and sequences on GenBank), suggesting that A29 (Collet *et al*. 2006) and 29a (Szalanski & Magnus 2010) are merely sequence variants of A16. The PCR–RFLP haplotype A16 has only been reported in the Portuguese territory, including mainland (Garnery *et al*. 1998a; Miguel *et al*. 2007) and islands (De la Rúa *et al*. 2006). Therefore, it is possible that haplotypes A29 and A29a descend from Portuguese colonies introduced by settlers in South America in historical times and later

expanded to North America (Szalanski & Magnus 2010) by Africanized honey bees. Alternatively, there were multiple independent, historical or recent, introductions in North and South America of putatively Portuguese-derived colonies.

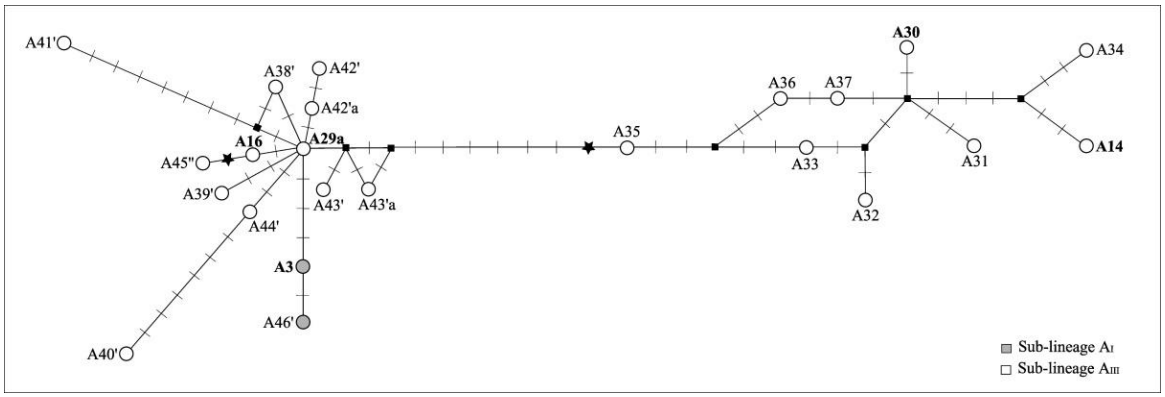


Figure III - 4. Median-joining network of the novel and previously described African haplotypes (marked in bold) that were identified from the tRNA^{leu}-cox2 region. Haplotypes are denoted as circles. Hypothetical (unsampled or extinct) haplotypes are symbolized as filled squares. The cross lines along branches represent mutational steps between nodes. The duplication of the Q element is represented by a star.

Similarities in primary and secondary structures between the Q element and the 3' end of COI gene (5' end of Q), tRNA^{leu} gene (middle part of Q), and the P sequence (3' end of Q) led Cornuet *et al*. (1991) to propose an origin of the Q element by tandem replication. The first and the second Q of haplotypes A41' and A40', respectively, bear a 3' end similar to P_1 whereas those of the other Q's are similar to P_0. Interestingly, the 3' end of Q1 of some lineage M haplotypes (M34-HQ337456.1; M43Q-HQ260365.1) and Z haplotypes (Z1-HM236204.1; Z12-HM236212.1; Z13-HM236213.1; Z1Q-HM236205.1) also bear a deletion motif similar to the P_1 element. While this deletion has probably multiple independent origins, this finding deserves further investigation as it may shed some light in the evolution of this complex region.

This study further expands on the complexity of the Iberian honey bee patterns and reinforces the importance of this southernmost European territory as a reservoir of *A. mellifera* genetic diversity. In response to a rapidly changing world (e.g., new pests and parasites, land use change), which has had severe consequences in apiculture, there is a growing alert for protecting honey bee genetic resources across its natural range (Jensen *et al*. 2005; De la Rúa *et al*. 2009; Dietemann *et al*. 2009; Haddad *et al*. 2009) and an increasing number of conservation programs, specially to protect *A. m. mellifera* (reviewed by De la Rúa *et al*. 2009). Preservation of honey bee genetic variation is a pre-requisite for long-term adaptive change and avoidance of fitness decline,

through inbreeding depression, and thereby a guarantee of a sustainable apiculture. The Iberian Peninsula has been a stage for evolutionary events that have shaped the evolutionary history of western European honey bee lineage. Therefore, this territory certainly deserves special attention in both small- and large-scale conservation programs.

References

Arias MC, Rinderer TE, Sheppard WS (2006) Further characterization of honey bees from the Iberian Peninsula by allozyme, morphometric and mtDNA haplotype analyses. *Journal of Apicultural Research*, **45**, 188–196.

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37–48.

Cánovas F, De la Rúa P, Serrano J, Galián J (2008) Geographical patterns of mitochondrial DNA variation in *Apis mellifera iberiensis* (Hymenoptera: Apidae). *Journal of Zoological Systematics and Evolutionary Research*, **46**, 24–30.

Cánovas F, De la Rúa P, Serrano J, Galián J (2011) Microsatellite variability reveals beekeeping influences on Iberian honeybee populations. *Apidologie*, **42**, 235–251.

Collet T, Ferreira KM, Arias MC, Soares AEE, Del Lama MA (2006) Genetic structure of Africanized honeybee populations (*Apis mellifera* L.) from Brazil and Uruguay viewed through mitochondrial DNA COI–COII patterns. *Heredity*, **97**, 329–335.

Cornuet JM, Fresnaye J (1989) Biometrical study of honey bee populations from Spain and Portugal. *Apidologie*, **20**, 93–101.

Cornuet JM, Garnery L, Solignac M (1991) Putative origin and function of the intergenic region between COI and COII of *Apis mellifera* L. mitochondrial DNA. *Genetics*, **128**, 393–403.

Crozier RH, Crozier YC (1993) The mitochondrial genome of the honey bee *Apis mellifera*: complete sequence and genome organization. *Genetics*, **133**, 97–117.

De la Rúa P, Serrano J, Galián J (1998) Mitochondrial DNA variability in the Canary islands honeybees (*Apis mellifera* L.). *Molecular Ecology*, **7**, 1543–1547.

De la Rúa P, Galián J, Serrano J, Moritz RFA (2001) Genetic structure and distinctness of *Apis mellifera* L. populations from the Canary Islands. *Molecular Ecology*, **10**, 1733–1742.

De la Rúa P, Galián J, Serrano J, Moritz RFA (2002) Microsatellite analysis of non-migratory colonies of *Apis mellifera iberica* from south-eastern Spain. *Journal of Zoological Systematics and Evolutionary Research*, **40**, 164–168.

De la Rúa P, Galián J, Serrano J, Moritz RFA (2003) Genetic structure of Balearic honeybee populations based on microsatellite polymorphism. *Genetics Selection Evolution*, **35**, 339–350.

De la Rúa P, Galián J, Pedersen BV, Serrano J (2006) Molecular characterization and population structure of *Apis mellifera* from Madeira and the Azores. *Apidologie*, **37**, 699–708.

De la Rúa P, Hernandez-Garcia R, Jiménez Y, Galián J, Serrano J (2005) Biodiversity of *Apis mellifera iberica* (Hymenoptera: Apidae) from northeastern Spain assessed by mitochondrial analysis. *Insect Systematics & Evolution*, **36**, 21–28.

De la Rúa P, Jaffé R, Dall'Olio R, Muñoz I, Serrano J (2009) Biodiversity, conservation and current threats to European honeybees. *Apidologie*, **40**, 263–284.

De la Rúa P, Jiménez Y, Galián J, Serrano J (2004) Evaluation of the biodiversity of honey bee (*Apis mellifera*) populations from eastern Spain. *Journal of Apicultural Research*, **43**, 162–166.

Dietemann V, Pirk CWW, Crewe R (2009) Is there a need for conservation of honeybees in Africa?. *Apidologie*, **40**, 285–295.

Engel MS (1999) The taxonomy of recent and fossil honey bees (Hymenoptera : Apidae; *Apis*). *Journal of Hymenoptera Research*, **8**, 165–196.

Franck P, Garnery L, Loiseau A *et al*. (2001) Genetic diversity of the honey bee in Africa: microsatellite and mitochondrial data. *Heredity*, **86**, 420–430.

Franck P, Garnery L, Solignac M, Cornuet JM (1998) The origin of west European subspecies of honeybees (*Apis mellifera*): new insights from microsatellite and mitochondrial data. *Evolution*, **52**, 1119–1134.

Garnery L, Franck P, Baudry E *et al*. (1998a) Genetic diversity of the west European honey bee (*Apis mellifera mellifera* and *A. m. iberica*). I. Mitochondrial DNA. *Genetics Selection Evolution*, **30**, S31–S47.

Garnery L, Franck P, Baudry E *et al*. (1998b) Genetic diversity of the west European honey bee (*Apis mellifera mellifera* and *A. m. iberica*). II. Microsatellite loci. *Genetics Selection Evolution*, **30**, S49–S74.

Garnery L, Mosshine EH, Oldroyd BP, Cornuet JM (1995) Mitochondrial DNA variation in Moroccan and Spanish honey bee populations. *Molecular Ecology*, **4**, 465–471.

Garnery L, Solignac M, Celebrano G, Cornuet JM (1993) A simple test using restricted PCR-amplified mitochondrial DNA to study the genetic structure of *Apis mellifera* L. *Experientia*, **49**, 1016–1021.

Gómez A, Lunt DH (2007) Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. In: Phylogeography of Southern European Refugia (eds Weiss S, Ferrand N), pp. 155–188. Springer, The Netherlands.

Haddad N, Meixner MD, Fuchs S *et al.* (2009) Mitochondrial DNA support for genetic reserves of *Apis mellifera syriaca* in Jordan. *Journal Apicultural Research*, **48**, 19–22.

Hewitt GM (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, **68**, 87–112.

Hewitt GM (2001) Speciation, hybrid zones and phylogeography—or seeing genes in space and time. *Molecular Ecology*, **10**, 537–549.

Jensen AB, Palmer KA, Boomsma JJ, Pedersen BV (2005) Varying degrees of *Apis mellifera ligustica* introgression in protected populations of the black honeybee, *Apis mellifera mellifera*, in northwest Europe. *Molecular Ecology*, **14**, 93–106.

Miguel I, Baylac M, Iriondo M *et al*. (2011) Both geometric morphometric and microsatellite data consistently support the differentiation of the *Apis mellifera* M evolutionary branch. *Apidologie*, **42**, 150–161.

Miguel I, Iriondo M, Garnery L, Sheppard WS, Estonba A (2007) Gene flow within the M evolutionary lineage of *Apis mellifera*: role of the Pyrenees, isolation by distance and postglacial re-colonization routes in the western Europe. *Apidologie*, **38**, 141–155.

Prada QCF, Duran JT, Salamanca GG, Del Lama MA (2009) Population genetics of *Apis mellifera* L. (Hymenoptera: Apidae) from Columbia. *Journal of Apicultural Research*, **48**, 3–10.

Rortais A, Arnold G, Alburaki M, Legout H, Garnery L (2011) Review of the DraI COI-COII test for the conservation of the black honeybee (*Apis mellifera mellifera*). *Conservation Genetics Resources*, **3**, 383–391.

Ruttner F, Tassencourt L, Louveaux J (1978) Biometrical–statistical analysis of the geographic variability of *Apis mellifera* L. *Apidologie*, **9**, 363–381.

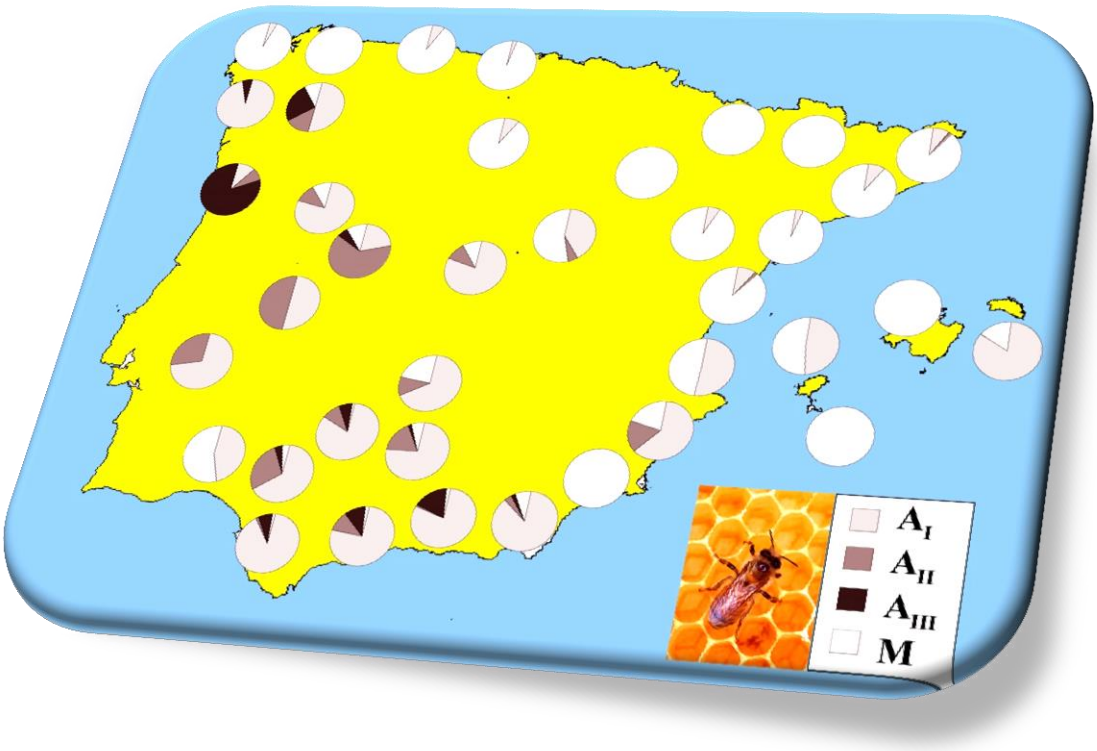
Sheppard, W.S., Meixner, M.D. (2003) *Apis mellifera pomonella*, a new honey bee subspecies from Central Asia. *Apidologie*. **34**, 367–375.

Smith DR, Glenn TC (1995) Allozyme polymorphisms in spanish honeybees (*Apis mellifera iberica*). *Journal of Heredity,* **86**, 12–16.

Smith DR, Palopoli MF, Taylor BR *et al*. (1991) Geographical overlap of two mitochondrial genomes in Spanish honeybees (*Apis mellifera iberica*). *Journal of Heredity*, **82**, 96–100.

Szalanski AL, Magnus RM (2010) Mitochondrial DNA characterization of Africanized honey bee (*Apis mellifera* L.) populations from the USA. *Journal of Apicicultural Research*, **49**, 177–185.

Tamura K, Peterson D, Peterson N *et al*. (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution*, **28**, 2731–2739.



Chapter IV

Revisiting maternal patterns of variation of *Apis mellifera iberiensis*: a revision based on sequence data of the tRNA^{leu}-cox2 mitochondrial intergenic region

Abstract

A revision of the Iberian honey bee diversity patterns, based on sequencing data of the tRNA^{leu}-cox2 mitochondrial DNA region, has refined the results of previous studies and demonstrated that these have missed a component of genetic variation when analysis is only based on the *Dra*I test. A description of haplotypes of Iberian honey bees has been updated based on nomenclature suggestions of Rortais *et al*. (2011). A fuller picture of the complex diversity of Iberian honey bees is revealed by the presence of 164 novel haplotypes, 113 belonging to the African lineage (A) and 51 to the Western European lineage (M). Within lineage A, 69 novel haplotypes belong to sub-lineage A$_{I}$, 13 to A$_{II}$, and 31 to A$_{III}$. For lineage M, two novel haplotypes show African and Western European features, based on sequence comparisons and phylogeny, suggesting that they could be ancestral haplotypes. A refinement of the previously reported Iberian cline formed by A and M haplotypes is further displayed. This study reinforces the Iberian Peninsula as an important source of honey bee maternal diversity and demands for further studies with evolutionary and conservation perspectives.

Keywords: Iberian honey bee, *Apis mellifera iberiensis*, tRNA^{leu}-cox2 intergenic region, ancestral haplotype M, *Dra*I test, sequence data

Introduction

The Western European honey bee, *Apis mellifera,* occupies a wide natural region that includes Europe, Africa, Western and Central Asia (Ruttner 1988). Its high dispersion capacity and ability to adapt to different ecological conditions led to emergence of 30 subspecies (Ruttner 1988, Engel 1999; Sheppard & Meixner 2003; Meixner *et al*. 2011). Morphological and molecular studies allowed the grouping of these numerous geographical subspecies into four major evolutionary lineages: African (A), Western and Northern European (M, hereafter referred to as Western European), Eastern European (C) and Western and Central Asia (O) (Ruttner 1988; Franck *et al*. 1998; Franck *et al.* 2001; Kandemir *et al*. 2011; Wallberg *et al*. 2014).

In Europe, the Iberian Peninsula has functioned as a glacial refuge for many species during cold periods of the Quaternary, accomplishing an important role in their evolutionary history (Taberlet *et al*. 1998; Hewitt 2000; Gómez & Lunt 2007). This area presents a wide range of climates, due to its geographical position (influenced by North Atlantic and Mediterranean), and physiographical complexity providing a variety of habitats and favoring the presence of multiple glacial refuges (Gómez & Lunt 2007). Furthermore, it has served as a contact zone between species from Europe and Africa during periods of expansion (Gibert *et al*. 2003; Molina-Venegas *et al*. 2015). These features have made Iberian Peninsula a place where processes of redistribution, fragmentation, contraction, isolation, expansion and admixture have occurred for many taxa during glacial and interglacial periods (Weiss & Ferrand 2007), and some of them have showed remarkable common patterns of phylogeographical concordance (Gómez & Lunt 2007), including the Iberian honey bee, *Apis mellifera iberiensis* (Chávez-Galarza *et al*. 2015).

The Iberian honey bee has been one of the most intensively surveyed honey bee subspecies in order to understand its complex pattern of diversity, typical of glacial refuges. Several phylogeographical studies have been performed based on morphology and behavior (Ruttner 1988; Cornuet & Fresnaye 1989; Arias *et al*. 2006; Miguel *et al*. 2011), pheromones (Hepburn & Radloff 1996), allozymes (Nielsen *et al*. 1994; Smith & Glenn 1995; Arias *et al*. 2006), mitochondrial DNA (Smith *et al*. 1991; Garnery *et al*. 1995, 1998a; De la Rúa *et al*. 2004a,b, 2005; Miguel *et al*. 2007; Cánovas *et al*. 2008; Pinto *et al*. 2012, 2013), microsatellites (Franck *et al*. 1998; Garnery *et al*. 1998b; Miguel *et al*. 2007, 2011; Cánovas *et al*. 2011) and Single Nucleotide Polymorphisms (Whitfield *et al*. 2006; Chávez-Galarza *et al*. 2013, 2015).

Numerous studies have used mitochondrial DNA (mtDNA), particularly the highly polymorphic tRNA^{leu}-cox2 intergenic region, which has allowed discrimination of four honey bee evolutionary lineages and has refined biogeographical patterns previously established from morphological data (Ruttner 1988; Garnery *et al*. 1995; Franck *et al*. 2001; Cánovas *et al*. 2008; Pinto *et al*. 2012, 2013).This region is characterized by the presence of a non-coding sequence formed by two elements named P and Q, both varying in size. In addition to distinguishing the four lineages, the P element allows discrimination of African sub-lineages (A_I, A_{II}, A_{III}, Z). The high content of adenine and thymine, stability profile, hairpin and cloverleaf putative secondary structures have suggested that this region would probably be a second putative origin of replication (Cornuet *et al*. 1991).

The *Dra*I test (Garnery *et al*. 1993), which consists of PCR-amplification of the tRNA^{leu}-cox2 region followed by digestion with the restriction enzyme *Dra*I, has been widely applied on Iberian honey bee genetic surveys (Garnery *et al*., 1995, 1998a; Franck *et al*. 1998, 2001; De la Rúa *et al*. 2004a,b, 2005; Miguel *et al*. 2007; Cánovas *et al*. 2008; Pinto *et al*. 2012, 2013). Analysis of mtDNA using this test has confirmed and refined the southwestern-northeastern cline formed by lineages A and M, as firstly reported by Smith *et al*. (1991). The maternal clinal pattern has been verified by a recent study using single nucleotide polymorphisms (SNPs) that characterized the whole genome of *A. m. iberiensis*, strengthening the hypothesis of ancestral secondary contact (Chávez-Galarza *et al*. 2015). Furthermore, the tRNA^{leu}-cox2 mtDNA region revealed high and complex haplotype diversity which is explained not only by the presence of haplotypes corresponding to lineage A and M but also by the geographical division of haplotypes from African lineage into three sub-lineages (A_I, A_{II} and A_{III}). Haplotypes belonging to sub-lineage A_I and A_{II} are mainly distributed in the Southern of Iberian Peninsula (Cánovas *et al*. 2008; Pinto *et al*. 2013) whereas haplotypes belonging to sub-lineage A_{III} are mostly restricted to the northern Atlantic side (Garnery *et al*. 1998a; Franck *et al.* 1998; Pinto *et al*. 2012, 2013), highlighting the great complexity of Iberian honey bees.

The aim of this study is to re-analyze the current status of haplotype diversity patterns in the Iberian honey bee, based on sequence data of the tRNA^{leu}-cox2 intergenic region. While maternal variation of Iberian honey bees has been extensively documented using restriction digestion data from the whole mitochondrial molecule (Smith *et al*. 1991; Garnery *et al*. 1992) or from the tRNA^{leu}-cox2 region (Franck *et al*. 1998, 2001; De la Rúa *et al*. 2004a, b, 2005; Miguel *et al*. 2007; Cánovas *et al*. 2008; Pinto *et al*. 2011, 2013), this is the first study reporting on

maternal diversity patterns using sequence data. By using sequence data, a more detailed picture of maternal patterns of Iberian honey bees will be provided as the *Dra*I test misses numerous haplotypes. Additionally, the sequence data will allow a revision of the haplotypes A and M as defined by *Dra*I band patterns.

Material and methods

Sampling

In 2010, a sampling was carried out across tree north-south transects in the Iberian Peninsula. One transect extended along the Atlantic coast (AT-Atlantic transect), one through the center (CT-central transect), and another along the Mediterranean coast (MT-Mediterranean transect). A total of 711 honey bee haploid males were collected, which represented 711 colonies (237 apiaries) grouped into 23 sites distributed along the three transects (AT=8, CT=9; MT=6) (see details in Chávez-Galarza *et al*. 2013). The geographical locations of individuals sampled are shown in Fig. IV - 1. To compare the patterns of genetic diversity of the Iberian honey bee with other subspecies, 31 males of *Apis mellifera intermissa* (lineage A), as well as the original sequence data previously obtained (see Pinto *et al*. 2014) for 34 males of *Apis mellifera mellifera* (lineage M), 17 individuals of *Apis mellifera ligustica* (Lineage C) and 19 individuals of *Apis mellifera carnica* (Lineage C) were also included in the analysis. Samples were collected from the inner part of hives, placed in absolute ethanol and stored at -20°C until molecular analysis.

Mitochondrial DNA extraction and sequencing

Total DNA was extracted from homogenized thorax of each male using the phenol-chloroform isoamyl alcohol (25:24:1) protocol (Sambrook *et al.* 1989). Analysis of mtDNA was based on the tRNA^{leu}-cox2 mitochondrial intergenic region. This region was PCR-amplified using the primer pair E2 (5'-GGC AGA ATA AGT GCA TTG-3') and H2 (5'-CAA TAT CAT TGA TGA CC-3'), according to a protocol detailed in Garnery *et al*. (1993). After PCR amplification, the products were sent to Macrogen (Seoul, Korea) for direct sequencing in both directions with primers E2 and H2. The sequences were checked for base calling and aligned with published sequences available in GenBank (http://www.ncbi.nlm.nih.gov/genbank) using MEGA 6.06 (Tamura *et al.* 2013).

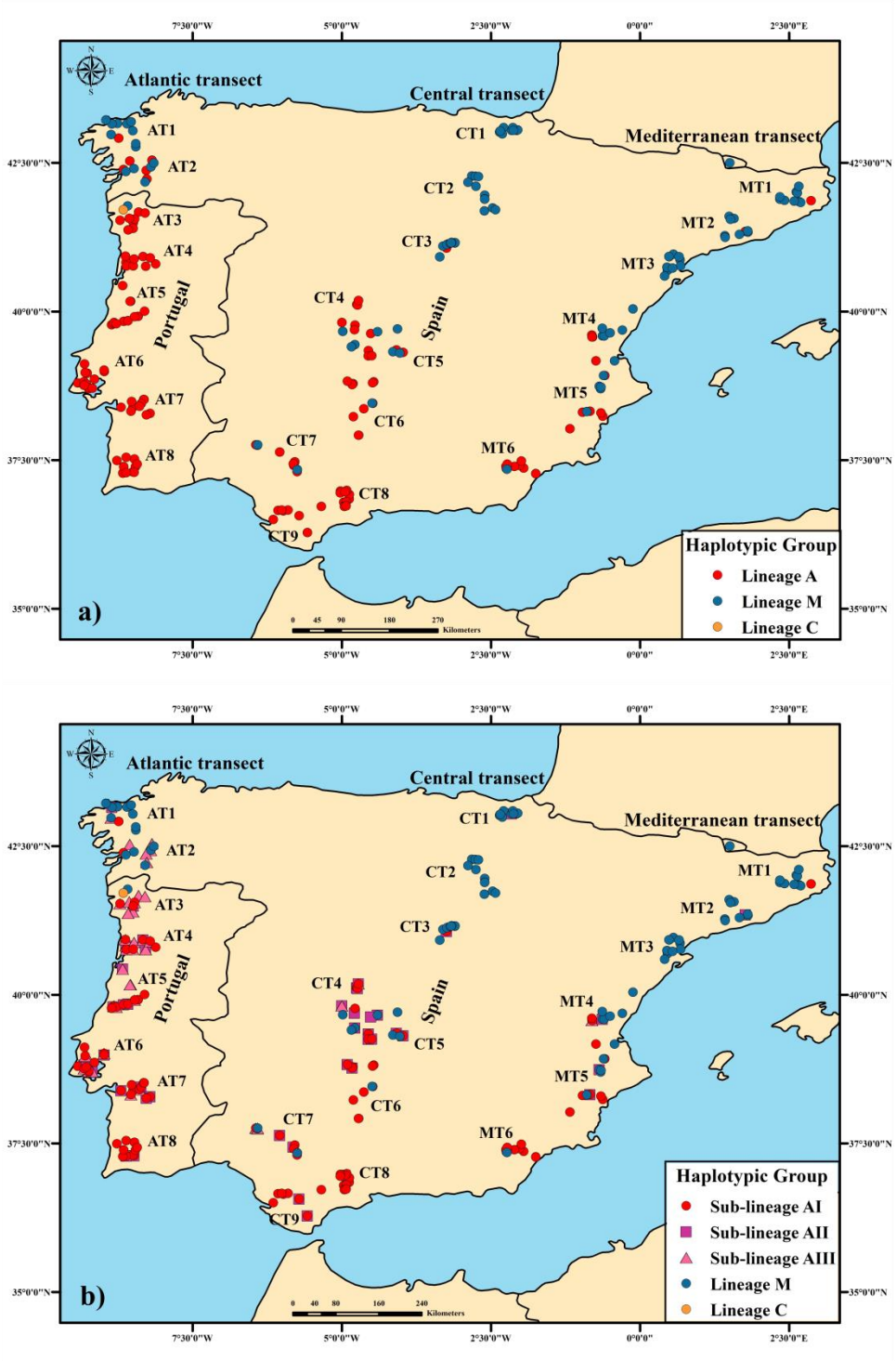


Figure IV - 1. Geographical location of sampled apiaries at each of the three transects (AT - Atlantic transect, CT - Central transect, MT - Mediterranean transect) in the Iberian Peninsula. Each point represents one colony. (a) Distribution of lineages, (b) distribution of lineages and African sub-lineages.

Structure of the tRNA^{leu}-cox2 intergenic region and nomenclature

The tRNA^{leu}-cox2 intergenic mtDNA region encompasses the 3' end of the tRNA^{leu} gene, the P and Q elements and the 5' end of the cox2 gene (Fig S IV - 1). This region has a variable size depending on the size and composition of the non-coding fragment formed by the P and Q elements. The P element ranges from ~53 to 68 bp whereas the Q element varies between ~194 and 196 bp. The P element exhibits three forms known as P_0, P, P_1. The Q element can be repeated in tandem one to five times, although the number of repeats is not lineage specific (Garnery *et al*. 1993; De la Rúa *et al*. 1998; Franck *et al*. 1998, 2001; Alburaki *et al*. 2011; Rortais *et al*. 2011). The Q element can be divided into three parts (Q_1, Q_2, and Q_3), which show a high level of similarity with the 3' end of the cox1 gene, the tRNA^{leu} gene and the P element, respectively.

Discrimination of the four lineages has been possible due to large deletions in the P element as well as restriction sites in the 3' end of tRNA^{leu} and the 5' end of the first Q element. Lineage A presents two forms of the P element: P_0 or P_1. While P_0 does not exhibit large deletions, P_1 is characterized by a large 15-bp deletion in the 3' end of P element (*d1* in Fig. Sup IV - 1). The form P_0 is found in sub-lineages A_I, A_{II} and Z whereas P_1 characterizes sub-lineage A_{III}. Furthermore, sub-lineage A_{II} is differentiated from sub-lineage A_I by the absence of the restriction site in the 5' end of the first Q element, and sub-lineage Z presents an additional restriction site in the middle of the first Q element. Lineage M is mainly distinguished by a large 13-bp deletion in the middle of the P element (*d* in Fig. Sup IV - 1) and two restriction sites in the first Q element like in the sub-lineage Z. The absence of the P element characterizes lineage C, which also exhibits a single Q element. In summary, the length of the intergenic region can be highly variable depending on the combination of number Q elements and forms of the P element.

Early studies using the *Dra*I test (PCR-RFLP) named the haplotypes using a upper-case letter corresponding to the evolutionary (A, M, C) and a number in order of appearance (Garnery *et al*. 1993, 1995). For example, haplotype A1 was the first African haplotype identified. The advent of sequencing revealed that the intergenic region is more complex than previously thought when considering only band patterns. Indeed, analysis of sequence data has allowed identification of nucleotide substitutions and small size indels undetected by the *Dra*I test. The increasing number of maternal surveys using the *Dra*I test has identified an increasing number of novel haplotypes, or sometimes just haplotype variants, which were frequently misnamed creating a great confusion when it comes to haplotype identification. In attempt to correct the situation, Rortais *et al*. (2011) proposed a standardization of nomenclature focusing on lineage M. The necessity of this

nomenclature arises because slight differences, as short indels and nucleotide substitutions could have been missed out. Rortais *et al*. (2011) proposed the following criteria: 1) haplotypes with the same *Dra*I test band pattern but different number of Q sequences must be labelled with the same number and added the symbols ', '', ''' to discriminate haplotypes with three ('), four (''), and five (''') Q elements, respectively, 2) haplotypes with one or two Q elements are differentiated by numbers, and 3) haplotypes with similar band patterns but slight variations (indels or substitutions) must be differentiated by lower-case letters (i.e. a, b, c, d, etc.).

In this study, the criteria of Rortais *et al*. (2011) were used, and two additional criteria were taken into account to label novel haplotypes. First, fragments between 27-65 bp with one nucleotide deletion or insertion, fragments between 66-110 bp with two nucleotides deletion or insertion, fragments between 111-200 bp with three nucleotides deletion or insertion, fragments between 201-300 bp with four nucleotides deletion or insertion, and fragments >301 bp with five nucleotides or more deletions or insertions must be assigned novel haplotypes, identified by a number. When the size of deletions or insertions was below those thresholds and in the presence of substitutions, the haplotypes were considered novel variants, identified by a lower-case letter. Second, when designating novel haplotypes their frequency is taken into account: the number is used for the most frequent haplotype whereas the letter is used for the least frequent.

Genetic diversity and phylogenetic analysis

The program GENEALEX 6.5 (Peakall & Smouse 2012) was used to obtain the following genetic diversity parameters within each sampling site: mean number of alleles per locus (N_a), effective number of alleles (N_e), number of private alleles (N_p), and unbiased diversity (*uh*). Genetic differentiation among Iberian sampling sites and among subspecies was estimated using Φ_{PT} values, which were then employed to perform a principal coordinate analysis (PCo) with the program GENEALEX and to build a neighbor-joining tree with the program POPULATIONS (Langella 2015).

Phylogenetic analysis of haplotypes was performed using the PHYLIP package 3.65c (Felsenstein 1993). The phylogenetic tree allowed identification of the lineages and sub-lineages to which the the novel haplotypes belong. Reference haplotypes from lineages A, M, and C obtained from NCBI (http://www.ncbi.nlm.nih.gov/) were included in the analysis (Franck *et al* 2001; Collet *et al*. 2006; Muñoz *et al*. 2009; Ödzil *et al*. 2009; Solórzano *et al*. 2009; Alburaki *et al*. 2011; Magnus *et al*. 2011; Muñoz and De la Rúa 2011; Rortais *et al*. 2011; Pinto *et al*. 2012, 2014;

Muñoz *et al*. 2013; Branchiccela *et al*. 2014; Techer *et al*. 2015). The analyses included polymorphic sites of the P element, the first Q element and the fragment of cox2. The remaining Q elements, which varied between two and three, were considered as single mutational steps. Gaps were considered as a fifth character. Positions indicating large deletions within the P element (P, P_1 and absence of P) and the first two restriction sites were upweighted. The tRNA^{leu}-cox2 region of *Apis cerana* was used as outgroup. The absence of the P element in lineage C and the absence of the Q element in *A. cerana* were coded as a single mutational step and the remaining positions as missing data.

Relationships among haplotypes and their proportions in the Iberian Peninsula were inferred using the median-joining network algorithm (Bandelt *et al*. 1999) as implemented in the program Network version 4.6.1.1 (Fluxus Engineering, Clare, UK; http://www.fluxus-engineering.com). The polymorphic sites within the P and Q elements were used to infer the network. Deletions from 2 to 35 bp were considered as a single-mutational step, being therefore coded as a 1-bp gap. Such as for phylogenetic analysis, positions indicating large deletions within the P element (P, P_1 and absence of P) and the first two restriction sites were also upweighted.

Pairwise t-tests between sampling sites were carried out to assess differences between the proportions of lineages and sub-lineages. The significant level was considered when *P*-value is below 0.05 (confidence interval of 95%). These analyses were performed using the R program (R Development Core Team 2013).

Results

Based on the tRNA^{leu}-cox2 intergenic region, haplotypes representing three different evolutionary lineages were detected among our samples: African (A), Western European (M) and Eastern European (C) (Table Sup IV - 1). Ten different length patterns (Q, P_0Q, P_0QQ, P_0QQQ, PQ, PQQ, PQQQ, P_1Q, P_1QQ, P_1QQQ), resulting from various combinations of the different P and Q elements, were observed among the 812 samples analyzed.

Revision of the names of haplotype

The GenBank reference haplotypes used in this study were reanalyzed and renamed following the criteria established by Rortais *et al*. (2011). Additionally, fragment sizes of haplotypes previously named from *Dra* I band patterns, for which there were no sequence data deposited in GenBank, were updated both in band pattern and size (Table Sup IV - 2). For example, haplotype

A3 has been described only as a *Dra*I band pattern, but there is no sequence data deposited in GenBank. The haplotypes A3, A10 and A16, which have three Q elements (Garnery *et al*. 1993, 1995; Franck *et al*. 1998), were renamed A3', A10' and A16'. The haplotypes A15 (De la Rúa *et al*. 1998) and A47 (Muñoz *et al*. 2013) exhibited band patterns similar to those originally proposed by Franck *et al*. (2001) for A14, although they differed in size due to the variable the number of Q elements. Hence, these haplotypes were renamed A14' (three Q elements) and A14'' (four Q elements), respectively. Sequence data for A4, reported more recently by Franck *et al*. (2001) and Collet *et al*. (2006), showed discrepancies with those firstly described by Garnery *et al*. (1993) and Franck *et al*. (1998). Specifically, while Franck *et al*. (2001) and Collet *et al*. (2006) reported band patterns of 47/107/191/483 and 47/108/192/483, respectively, those reported by Garnery *et al*. (1993) and Franck *et al*. (1998) were 47/108/193/483. Therefore, haplotype A4 identified by Collet *et al*. (2006) was maintained, because it is more similar to the original proposal, while that of Franck *et al*. (2001) was renamed as A4a. Similarly, haplotype A1 and its variants deposited in GenBank, were not concordant with the original description (Garnery *et al*. 1993, Garnery *et al*. 1995). The band pattern of A1b proposed by Franck *et al*. (2001) should be registered as A1 because it exhibits the original features firstly described by Garnery *et al*. (1993) and Franck *et al.* (1998). Sequences reported by Collet *et al*. (2006) and Branchiccela *et al*. (2014) showed differences in size and nucleotide substitutions with the original proposal, being reassigned as A1b and A1e, respectively.

Such as for lineage A, we renamed several haplotypes from lineage M (Table Sup IV - 2). Haplotype M4 has been reported in various studies (Garnery *et al*. 1993; Franck *et al*. 1998, 2001; Collet *et al*. 2006; Rortais *et al*. 2011; Pinto *et al*. 2014), although sequence data were only deposited in GenBank by Franck *et al*. (2001), Rortais *et al*. (2011), and Pinto *et al*. (2014). While the sequences of Franck *et al*. (2001) and Pinto *et al*. (2014) are very similar (only differing in a C/T transition at the 5' end of cox2), and match the original *Dra*I band pattern (142/65[2]/131/422), the sequence deposited by Rortais *et al*. (2011) is very different. Considering the low frequency of the sequence reported by Franck et al. (2001) and the high frequency found by Pinto et al. (2014) and this study (Table Sup IV - 3), we propose to rename the M4a of Pinto et al. (2014) as M4 and the M4 of Franck et al. (2001) as M4a. The M4 reported by Collet *et al*. (2006) is similar to the M17 of Rortais *et al*. (2011), and should therefore be renamed M17b. Some M4 variants named by Pinto *et al*. (2014) were re-analyzed and identified and renamed as M17 variants (M17b to M17f, M17h), although the M4i of Pinto *et al*. (2014) exhibits features of a

new haplotype and was therefore assigned a new haplotype named M71 (Table Sup IV - 2). While Rortais *et al*. (2011) proposed a supposedly final description of M4', the sequence data uploaded in GeneBank is not congruent with the *Dra*I band pattern of M4'. Therefore, the correct sequence is presented in this study. Finally, we detected incongruences in fragment sizes and band patterns reported for haplotypes M10, M11, M12, M13, M34, M55', M62, and M63, which were corrected according to the sequence data deposited in GenBank (see Table Sup IV - 2).

Although haplotypes of C ancestry are very rare in the Iberian Peninsula (Cánovas *et al*. 2008, Chávez-Galarza *et al*. 2015), we also propose a revision for C haplotypes showing incongruences. While haplotype C2 (Franck *et al*. 2001) and C3 (Perrier *et al*. 2003) have been reported, their sequences are not available in GenBank. The haplotypes C2 and C11 reported by Techer *et al*. (2015) and Solórzano *et al*. (2009), respectively, should be reassigned to C2j (Muñoz & De la Rúa 2011) because the three sequences are perfect matches. The haplotype C2d was reassigned C2 because of its high frequency (Muñoz *et al*. 2009). The band patterns of haplotypes C2e (Muñoz *et al*. 2009), C2k (Razpet *et al*. unpublished) and C31 (Magnus *et al*. 2011) were similar to those of C3 and were therefore renamed C3, C3a, C3b, respectively. Finally, the sequence of isolate C34 (Magnus *et al*. 2011) was assigned C34.

Phylogenetic analysis

Phylogenetic analysis of the novel haplotypes is shown in Fig. IV - 2. The novel haplotypes identified in Iberian honey bees and reference subspecies, were grouped within lineages A (131), and M (51) (Table Sup IV - 1). In lineage A, 72 haplotypes belong to sub-lineage A_{I}, 28 to sub-lineage A_{II}, 31 to sub-lineage A_{III}, and none was identified as sub-lineage Z. Lineage A and M formed a group, with sub-lineage Z as a basal group. Sub-lineages A_{I}, A_{II} and A_{III} are grouped within one clade, although A_{I} is not differentiated (Fig. IV - 2). The phylogeny supported three African sub-lineages forming a sister group of lineage M. Lineage M formed a well-differentiated group, with novel haplotypes M79 and M79a forming a new subgroup. The main distinctive feature of these haplotypes is to possess a P element similar to P_{0}, which is typical of A haplotypes, although they exhibit nucleotide substitutions and restriction sites characteristic of M haplotypes (Fig. Sup IV - 1). The group formed by haplotypes of C ancestry was clearly differentiated from the remaining haplotypes.

Distribution of haplotypes

In the Iberian Peninsula, a total of 188 different haplotypes was observed among the 711 sampled colonies of which 128 belong to lineage A, 59 to lineage M and 1 to lineage C (Table Sup IV - 1). Within lineage A, 74 haplotypes belong to sub-lineage A_{I}, 15 to sub-lineage A_{II}, and 39 to sub-lineage A_{III}. By comparison with the GenBank sequences, we characterized 24 previously reported haplotypes (Franck *et al.* 2001), and 164 (113 A and 51 M) novel haplotypes (Table Sup IV - 1). Lineage A was dominant (481 colonies, 67.5%) as compared with M (229 colonies, 32.2%) and C (1 colony, 0.3%) lineages. Within lineage A, sub-lineage A_{I} (333 colonies, 69.2%) was also more ubiquitous than A_{II} (68 colonies, 14.2%) and A_{III} (80 colonies, 16.6%).

The reference subspecies *A. m. ligustica* presented a mixture of haplotypes M (5 colonies, 29.4%) and C (12 colonies, 70.6%), and *A. m. intermissa*, *A. m. mellifera* and *A. m. carnica* carried haplotypes of a single lineage. *A. m. mellifera* presented haplotypes M (34 colonies, 100%), and *A. m. carnica* presented haplotypes C (19 colonies, 100%). Only *A. m. intermissa* presented 18 novel African haplotypes: 3 of sub-lineage A_{I} (5 colonies, 16.1%) and 15 sub-lineage A_{II} (26 colonies, 83.9%).

As previously reported, haplotype A2 is the most frequent African haplotype (6.6 %), occurring in almost all areas where the African lineage is found, followed by A2a (6.5%), A1 (5.9 %), A8 (5.6 %) and A1h (3.8 %). Among haplotypes of M origin, M4 is the most frequent (9.7 %), occurring mostly in the Northeastern part of Iberia, followed by M7 and M7n with 2.1 %. All these haplotypes were observed in more than 7 sampling sites (Table Sup IV - 3). While some haplotypes are distributed across Iberia, some are more concentrated in certain areas, as follows: M7 and A49 (AT1), A16' (AT2), A16b' (AT3), A1h (AT4), A20 (AT5), A1 (AT6 to AT8), M4 (CT1, CT2), M79 (CT3), A8 (CT4, CT5), A2 (CT6, CT7, CT9), A2 and A2c (CT8), M4 (MT1 to MT3), A2a (MT4 to MT6) (Table Sup IV - 3).

Relationships among haplotypes

A median-joining network that illustrates the frequencies and relationships among the haplotypes found in the Iberian Peninsula is shown in Fig. IV - 3. Two highly divergent clusters, corresponding to lineages A and M, are identified. For lineage A, the three African sub-lineages were mainly linked by haplotypes A1 and A2 (sub-lineage A_{I}), A8 and A9a (sub-lineage A_{II}), and A11, A20 and A30 (sub-lineage A_{III}). Haplotypes A1, A2, A8 and A16' exhibit the greatest number of links with other haplotypes. For lineage M, haplotypes M79 and M79a present an almost intermediate position

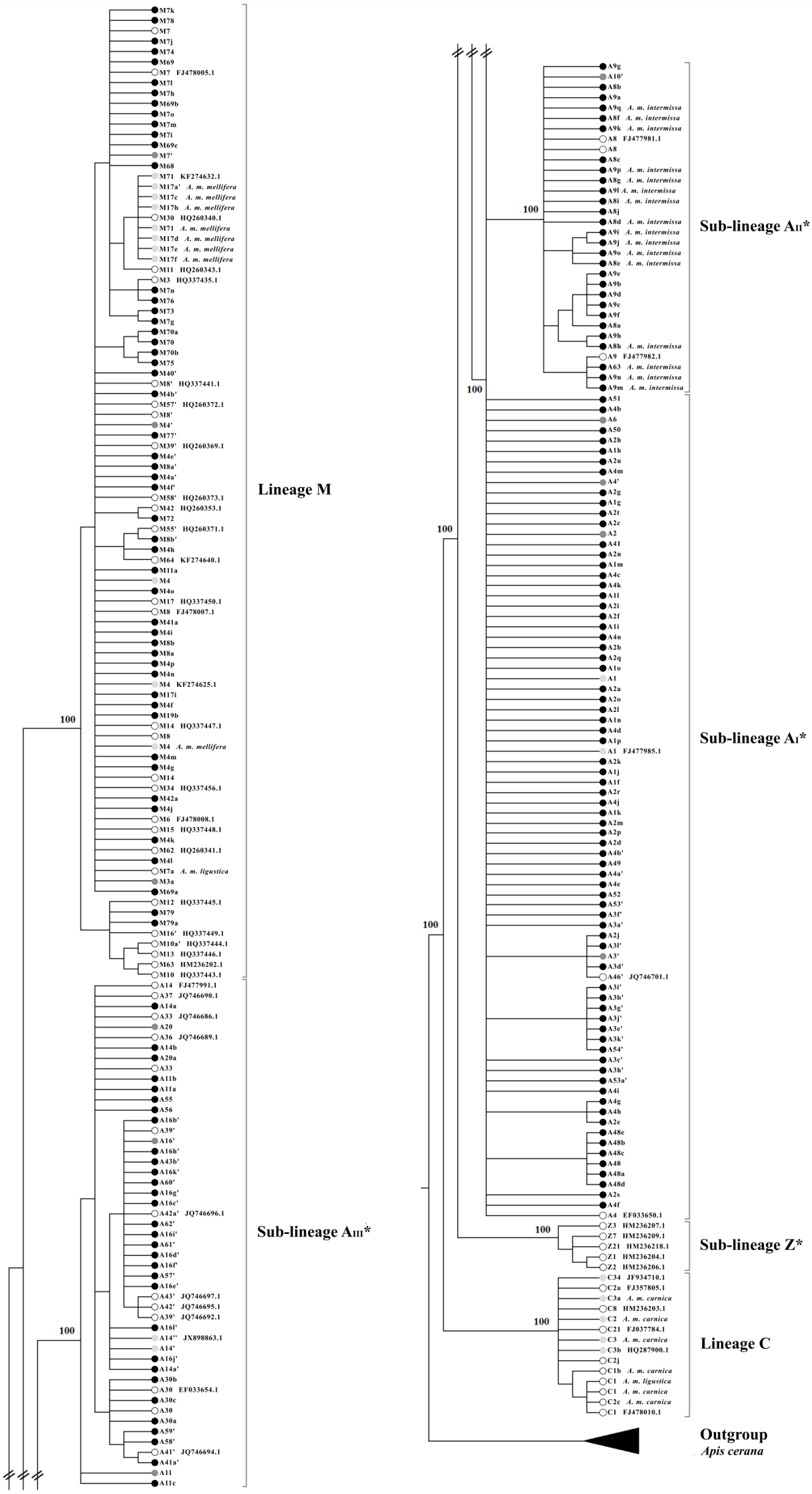


Figure IV - 2. Phylogeny of haplotypes based on the tRNA^{leu} -cox2 intergenic region for Iberian (*A. m. iberiensis*) and reference (*A. m. mellifera*, *A. m. ligustica*, and *A. m. carnica*) honey bee subspecies. Strict consensus (level 95%) from over 10 000 equally parsimonious trees. Values indicate bootstrap support of 1000 pseudoreplicates. *Sub-lineages A_I, A_{II}, A_{III}, and Z belong to the African lineage. Black circles indicate novel haplotypes, grey circles indicate haplotypes (band patterns) reported by others but sequenced here for the first time, light grey circles indicate renamed haplotypes, and white circles indicate haplotypes reported by others with a complete description (*Dra*I band patterns and sequence data)

between haplotypes M and A, as observed in the phylogenetic tree. Haplotypes M4 and M7 occupy a central position between two divergent groups and show more links than the other M haplotypes.

Diversity measures

Diversity measures estimated for each Iberian sampling site and reference population is shown in Table IV - 1. The number of haplotypes (Na) ranged from 7 in AT8 to 20 in AT4. The sampling sites AT1 and AT3 presented the highest value of effective number of alleles (Ne). For the number of private alleles (Np), the highest value was observed in AT4 followed by AT3, AT1 and AT6. The unbiased haplotype diversity (*uh*) ranged from 0.956 (AT1, AT3) to 0.464 (AT8). Globally, the highest diversity estimates were found across the Atlantic transect (except for AT8). When comparisons were made with reference populations, *A. m. intermissa* showed the highest diversity values and *A. m. ligustica* the lowest.

Relationships among lineages

The neighbor-joining tree of Iberian sampling sites and reference subspecies showed three well-supported groups (Fig. IV - 4a). The first group is formed by populations with high proportion of lineage C haplotypes (*A. m. ligustica* and *A. m. carnica*). The second group includes mostly M haplotypes of the reference subspecies *A. m. mellifera* and *A. m. iberiensis* from sampling sites AT1, CT1 to CT3, MT1 to MT3. The third group is comprised by the reference *A. m. intermissa* and *A. m. iberiensis* from sampling sites AT2 to AT8, CT4 to CT9, MT4 to MT6. Within the third group, populations of CT4 and CT5 and *A. m. intermissa* formed a well-supported group dominated by sub-lineage A_{II}. Sub-lineage A_{III} occurred in high proportion in AT3 and AT4, while sub-lineage A_{I} was the most frequent in AT7 to AT8, CT6 to CT9, and MT5 to MT6.

Similarly, analysis of PCo grouped reference subspecies and *A. m. iberiensis* sampling sites as the neighbor-joining tree (Fig. IV - 4b). The main axis explained 78.5% of the genetic variation, separating the three aforementioned groups. The second axis explained 14.7% of the genetic variation and allowed a better differentiation among African sub-lineages and lineage C.

Discussion

The analysis using sequence data of the tRNA^{leu}-cox2 region has revealed a total of 164 novel haplotypes belonging to African (113) and Western European (51) lineages. This finding represents

a 188.3% and 46.8% increase of documented A and M haplotypes, respectively, taking into account the 60 African and 109 Western European available haplotypes (Garnery *et al*. 1993, 1995; De la Rúa *et al*. 1998, 2005; Franck *et al*. 2001; Collet *et al*. 2006; Magnus and Szalanski 2010; Rortais *et al*. 2011; Pinto *et al*. 2012, 2014; Muñoz *et al*. 2013; Bertrand *et al*. 2015). Within lineage A, a greater number of novel haplotypes were detected in sub-lineage A_{I} (69) than in sub-lineages A_{II} (13) and A_{III} (31), representing an increase of 246.4%, 216.6% and 119.2%, respectively. This finding re-emphasizes the Iberian Peninsula as an important source of honey bee maternal diversity (Franck *et al* 1998; Cánovas *et al*. 2008; Pinto *et al*. 2012), and reveals the usefulness of sequencing the $tRNA^{leu}$-cox2 intergenic region for revealing a more detailed picture of the maternal diversity of the Iberian honey bee.

The *Dra*I test has proved very effective in determining the maternal origin of honey bee colonies, which explains its popularity revealed by the numerous publications using this method (see references in Rortais *et al*. 2011). However, short indels in large fragments and nucleotide substitutions are hardly distinguished in acrylamide or high resolution agarose gels (Rortais *et al*. 2011). Therefore, sequencing is recommended and would help unraveling, describing, and resolving cryptic diversity patterns, as shown here. Previous studies based on *Dra*I test indicated A2 as the most frequent haplotype in the Iberian Peninsula, which was mainly located in the South of Spain (Franck *et al*. 1998; Miguel *et al*. 2007; Cánovas *et al*. 2008). However our sequence data highlighted the presence of two haplotypes with similar frequencies characterizing the South of Spain: A2 and A2a. This finding indicates an overestimation of haplotype A2 frequency, as determined by the *Dra*I test. Haplotype A2 predominates in the South of the central transect, and A2a in the South of the Mediterranean transect (see Table Sup IV - 3). Although, A2a is almost as abundant as A2, the latter exhibits a central position in the network and displays more connections with other haplotypes (Fig. IV - 3), which suggests that it is an ancestral haplotype. Our results also suggest that haplotype A8 is an ancestral haplotype due to its high frequency, central position and several connections for sub-lineage A_{II}. Furthermore, the presence of haplotype A30 in the Iberian Peninsula has confirmed the usefulness of sequencing the $tRNA^{leu}$-cox2 region to trace the origin of haplotypes found in Brazil (Collet *et al*. 2006), which were most likely introduced by Portuguese settlers.

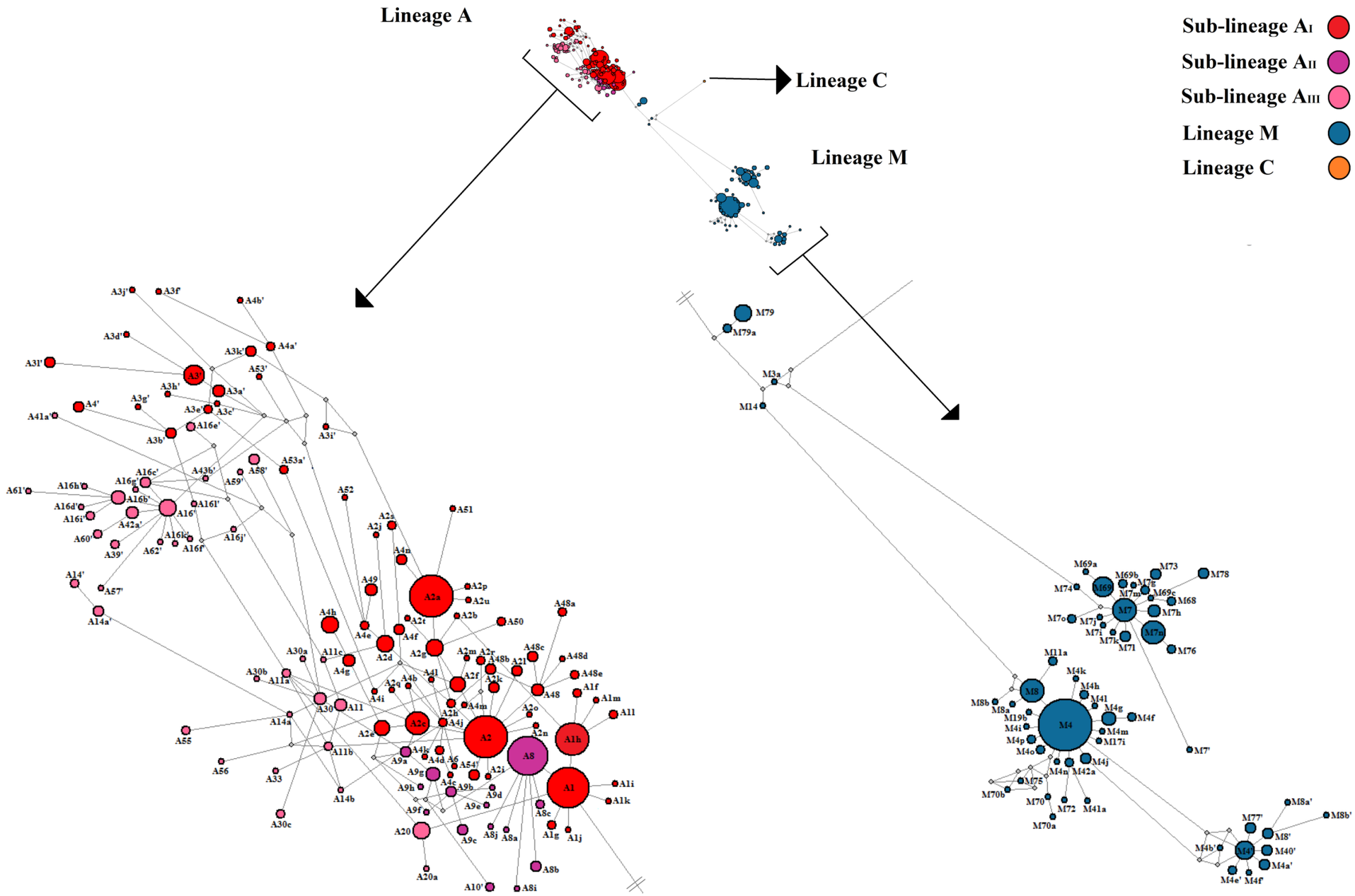


Figure IV - 3. Median-joining network of haplotypes identified in the Iberian Peninsula. Hypothetical (unsampled or extinct) haplotypes are indicated as grey filled squares. The size of circles is proportional to the haplotype frequencies. Links between haplotypes are proportional to genetic distances between them.

Table IV - 1. Diversity measures for each of the 23 sampling sites and reference subspecies. Sampling sites are indicated as AT1 - MT6 (AT - Atlantic transect, CT - Central transect, MT - Mediterranean transect) and reference subspecies as INT-MEL-LIG-CAR (INT - *A. m. intermissa*, MEL - *A. m. mellifera*, LIG - *A. m. ligustica*, CAR - *A. m. carnica*). N -number of individuals, Na - mean number of alleles, Np - number of private alleles, Ne - number of effective alleles and *uh* - unbiased haplotype diversity.

Location	N	Na	Np	Ne	*uh*
Sampling sites					
AT1	30	18	11	13.235	0.956
AT2	30	16	8	10.714	0.938
AT3	30	18	13	13.235	0.956
AT4	30	20	16	12.857	0.954
AT5	33	14	5	8.712	0.913
AT6	39	18	10	12.168	0.942
AT7	30	14	6	9.783	0.929
AT8	30	7	1	1.815	0.464
CT1	30	11	5	6.618	0.878
CT2	30	8	2	3.913	0.770
CT3	30	12	8	7.377	0.894
CT4	30	10	2	4.545	0.807
CT5	30	13	3	7.627	0.899
CT6	33	13	4	7.949	0.902
CT7	30	16	4	12.500	0.952
CT8	30	13	4	6.000	0.862
CT9	24	11	4	4.721	0.822
MT1	36	14	8	7.624	0.894
MT2	30	13	6	5.556	0.848
MT3	30	16	9	7.627	0.899
MT4	33	13	5	5.261	0.835
MT5	33	16	8	6.444	0.871
MT6	30	8	3	4.737	0.816
Reference subspecies					
INT	31	18	18	14.343	0.961
MEL	34	8	7	3.803	0.759
LIG	17	2	1	1.710	0.441
CAR	19	6	5	3.374	0.743

The haplotypes form a well-defined cline in Iberia with haplotypes of A ancestry predominating in the Southwestern half and haplotypes of M ancestry predominating in the Northeastern half (Fig. IV - 1a). Previous mitochondrial studies have reported this clinal pattern (Smith *et al*. 1991; Garnery *et al*. 1992, 1998a; Cánovas *et al*. 2008), which has been supported by nuclear markers as SNPs (Chávez-Galarza *et al*. 2015). Taken together, these results rescue and reinforce the hypothesis of a hybrid origin for *A. m. iberiensis* resulting from the coexistence of individuals of lineage M and A through a process of secondary contact (Smith *et al*. 1991; Garnery *et al*. 1992; Chávez-Galarza *et al*. 2015). The very low frequency of colonies carrying C-lineage

haplotypes (one individual among 711) has seemingly no impact on the observed clinal pattern suggesting that the influence of humans has not currently affected the biogeographical structure of the Iberian honey bee.

Populations of honey bees have probably entered Iberia through the Strait of Gibraltar (Ruttner *et al*. 1978; Whitfield *et al*. 2006; Han *et al*. 2012), from an origin in either Africa (Whitfield *et al*. 2006) or Western Asia (Wallberg *et al*. 2014). Several studies have reported a large frequency of individuals with haplotypes belonging to sub-lineage A_{I} distributed in Africa, Mediterranean Basin and Iberia (Garnery *et al*. 1993, 1995; De la Rúa *et al*. 2001; Franck *et al*. 2000, 2001; Cánovas *et al*. 2008), being confirmed in this study for the latter (Fig. IV - 1b). These results have led to suggestion that individuals belonging to sub-lineage A_{I} have been the first colonizers, followed by sub-lineage A_{II} and later on by sub-lineage A_{III} (Cánovas *et al*. 2008). While haplotypes of sub-lineage A_{II} ancestry have been documented as the least frequent (Cánovas *et al*. 2008), we found a high frequency of these haplotypes in the central part of Spain (Fig. IV - 1b, Fig. IV - 4). Our results further confirmed the high maternal diversity reported for sub-lineage A_{III} (Fig. IV - 3), and its Ibero-Atlantic distribution (Pinto *et al.* 2012, 2013), suggesting that an independent wave might have expanded into the northwestern part of Iberia (Cánovas *et al*. 2008), or, alternatively, might have evolved in this area.

The complex haplotype network of haplotypes A and M shown in Fig. IV - 4 for the Iberian Peninsula might be explained by multiple origins of haplotypes (Crandall & Templeton 1993), and by a more recent history of diversification for haplotypes belonging to African sub-lineages. The links with large genetic distance that separate the most frequent haplotypes in lineage M represent several mutational steps that might have accumulated during the climatic oscillations produced in the glacial episodes. Conversely, lineage A exhibits shorter links than lineage M for the most frequent haplotypes, suggesting that individuals of African ancestry could have been affected in less extent by the climatic changes of the Pleistocene and developing in a more stable climate. Nonetheless, during cold periods, lineage M might have used a refuge an area located between Iberian Mountain Range and Pyrenees for (Miguel *et al*. 2007), and lineage A an area in the Betic Ranges of Southern Spain (Chávez-Galarza *et al*. 2015), as suggested for other taxa (Gómez & Lunt 2007). The putative refuge in the Betic Ranges might have mostly harbored honey bees of A_{I} and A_{II} ancestry, but the more confined distribution of sub-lineage A_{III} to the North Atlantic side suggests a putative third refuge. Putative Atlantic refuge has been proposed for thermophilous flora (González-

Sampériz *et al*. 2010), which is of well-known importance for the development of honey bee colonies and might have been crucial for their survival during the glaciations.

The origin of honey bee evolutionary lineages has been attempted to be explained from morphological data (Ruttner 1988), allozymes (Smith & Glenn 1995), mtDNA (Garnery *et al*. 1992), and SNPs (Whitfield *et al*. 2006; Han *et al*. 2012). Two hypotheses have been referred about the origin of lineage M. The first one speculates that lineage M has derived from African honey bee populations (Ruttner 1988; Whitfield *et al*. 2006; Han *et al*. 2012). The second one suggests that both lineages M and A have diverged in the origin center of genus *Apis* (Garnery *et al*. 1992). Although, morphological and allozyme data more clearly detect Iberian individuals with intermediate A and M characteristics, this study reports for the first time haplotypes (M79, M79a) with these particularities (Fig. Sup IV - 1), referring them as kind of 'ancestral' M haplotypes. While, our results highlight an increase in diversity for lineage M, those do not completely resolve its putative origin. According to the topology shown in Fig. IV - 2, sub-lineage Z is seemingly an ancestral group from which the three African sub-lineages and lineage M would have diverged. A similar result was found by Alburaki *et al*. (2011) using Cytb sequences, which enabled the authors determining the relationship of sub-lineage Z inside the African lineage. Within lineage M, the group formed by 'ancestral' haplotypes M formed a well-defined group other than the typical haplotypes M, suggesting a clear divergence among them (Fig. IV - 2). This finding suggests that lineage M would have a more ancestral African origin, not from North African populations but from Northeastern African and Near East populations. This conjecture is based on the presence of haplotypes Z in *A. m. lamarckii* and *A. m. syriaca* from Egypt, Syria and Lebanon (Alburaki *et al*. 2011). This observation deserves however further research using phylogenomics since phylogeny applied to the intergenic region has mainly served for identifying haplotypes within lineages or sub-lineages, and not for inferring divergence time due to the presence of many gaps/indels.

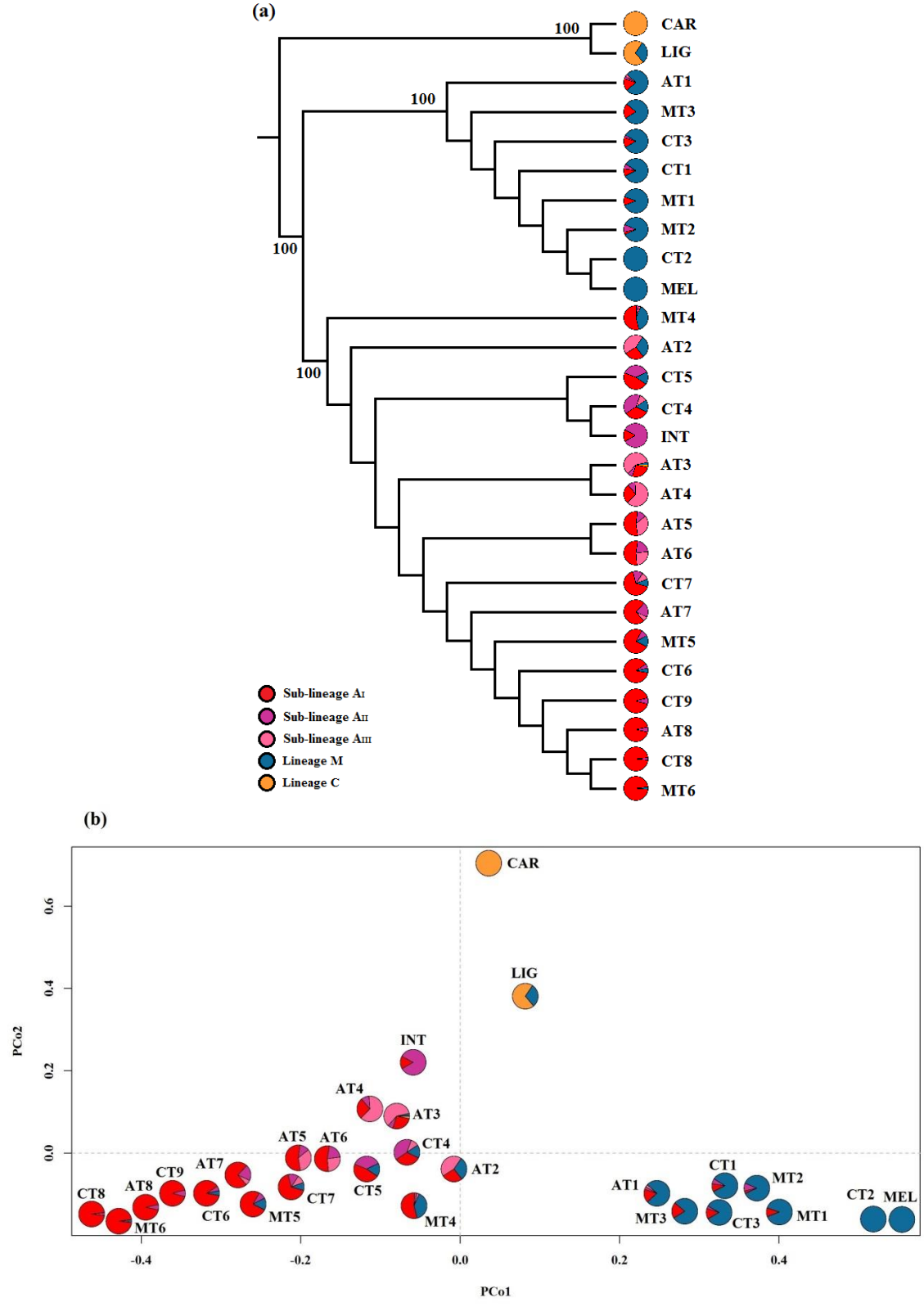


Figure IV - 4. Genetic relationships between sampling sites and reference subspecies using Φ_{PT} values obtained with frequency data for lineages and sub-lineages. Pie charts display the proportion between lineages and sub-lineages. (a) Neighbor-joining tree with support of 1000 bootstraps. (b) PCo presenting 78.5% and 14.7% of total variation for each axis, respectively.

This study further expands our knowledge on the complex diversity of tRNA^{leu}-cox2 in the Iberian Peninsula and reinforces the utility of sequencing as a complement to the *Dra*I test. In accordance with our results, we suggest the use of gene genealogies and/or genomes to elucidate the putative presence of a third refuge in North Atlantic side as it has been proposed for other organisms (González-Sampériz *et al*. 2010; Gómez & Lunt 2007), and to explain the origin of evolutionary lineages and resolve the incongruences revealed by the different markers. It is necessary to include paleontological and paleoclimatological data to provide record on the past honey bee distribution. An important event that has never been taken into account is known as 'Messinian crisis salinity', which occurred in late Miocene (approximately 7.2 – 5.3 Mya) when *Apis mellifera* diverged from *A. cerana* (Garnery *et al*. 1991). This event, originated by desiccation of Mediterranean Sea, formed a large land bridge for many animals between Africa and the Iberian Peninsula (Gibert *et al*. 2013) and might have enabled early honey bee colonization. The paleontological records should be combined with genomic data for estimating divergence time and diversity center. On the other hand, human-assisted processes (transhumance, introduction of exotic queen, *etc*) have not been shocking altering the clinal pattern. In spite of high dispersion abilities in honey bee, the clinal pattern has been maintained which could also be an effect of natural selection (Cánovas *et al*. 2014). Understanding the history and underlying processes shaping Iberian honey bee diversity would be an important first step when initializing conservation programs with view to sustainable apiculture and regional development.

References

Alburaki M, Moulin S, Legout H, Alburaki A, Garnery L (2011) Mitochondrial structure of Eastern honeybee populations from Syria, Lebanon and Iraq. *Apidologie*, **42**, 628–641.

Arias MC, Rinderer TE, Sheppard WS (2006) Further characterization of honey bees from the Iberian Peninsula by allozyme, morphometric and mtDNA haplotype analyses. *Journal of Apicultural Research*, **45**, 188–196.

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37–48.

Bertrand B, Alburaki M, Legout H *et al*. (2015) MtDNA COI-COII marker and drone congregation area: An efficient method to establish and monitor honeybee (*Apis mellifera* L.) conservation centres. *Molecular Ecology Resources*, **15**, 673-683.

Branchiccela B, Aguirre C, Parra G *et al*. (2014) Genetic changes in *Apis mellifera* after 40 years of Africanization. *Apidologie*, **45**, 752-756.

Cánovas F, De la Rúa P, Serrano J, Galián J (2008) Geographical patterns of mitochondrial DNA variation in *Apis mellifera iberiensis* (Hymenoptera: Apidae). *Journal of Zoological Systematics and Evolutionary Research*, **46**, 24–30.

Cánovas F, De la Rúa P, Serrano J, Galián J (2011) Microsatellite variability reveals beekeeping influences on Iberian honeybee populations. *Apidologie*, **42**, 235–251.

Cánovas F, De la Rúa P, Serrano J, Galián J (2014) Analysis of a contact area between two distinct evolutionary honeybee units: an ecological perspective. *Journal of Insect Conservation*, **18**, 927–937.

Chávez-Galarza J, Henriques D, Johnston JS *et al*. (2013) Signatures of selection in the Iberian honey bee (*Apis mellifera iberiensis*) revealed by a genome scan analysis of single nucleotide polymorphisms. *Molecular Ecology*, **22**, 5890–5907.

Chávez-Galarza J, Henriques D, Johnston JS *et al*. (2015) Revisiting the Iberian honey bee (*Apis mellifera iberiensis*) contact zone: maternal and genome-wide nuclear variations provide support for secondary contact from historical refugia. *Molecular Ecology*, **24**, 2973–2992.

Collet T, Ferreira KM, Arias MC, Soares AEE, Del Lama MA. (2006) Genetic structure of Africanized honeybee populations (*Apis mellifera* L.) from Brazil and Uruguay viewed through mitochondrial DNA COI–COII patterns. *Heredity*, **97**, 329–335.

Cornuet JM, Fresnaye J (1989) Biometrical study of honey bee populations from Spain and Portugal. *Apidologie*, **20**, 93–101.

Cornuet JM, Garnery L, Solignac M (1991) Putative origin and function of the intergenic region between COI and COII of *Apis mellifera* L. mitochondrial DNA. *Genetics*, **128**, 393–403.

Crandall KA, Templeton AR (1993) Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. Genetics, 134, 959-969.

De la Rúa P, Galián J, Serrano J, Moritz RFA (2001) Molecular characterization and population structure of the honeybees from the Balearic islands (Spain). *Apidologie*, **32**, 417–427.

De la Rúa P, Hernandez-Garcia R, Jiménez Y, Galián J, Serrano J (2005) Biodiversity of *Apis mellifera iberica* (Hymenoptera: Apidae) from northeastern Spain assessed by mitochondrial analysis. *Insect Systematics & Evolution*, **36**, 21–28.

De la Rúa P, Hernandez-Garcia R, Pedersen BV, Galián J, Serrano J (2004a) Biodiversity of *Apis mellifera iberica* (Hymenoptera: Apidae) from northeastern Spain assessed by mitochondrial analysis. *Insect Systematics & Evolution*, **36**, 21–28.

De la Rúa P, Jiménez Y, Galián J, Serrano J (2004b) Evaluation of the biodiversity of honey bee (*Apis mellifera*) populations from eastern Spain. *Journal of Apicultural Research*, **43**, 162–166.

De la Rúa P, Serrano J, Galián J (1998) Mitochondrial DNA variability in the Canary islands honeybees (*Apis mellifera* L.). *Mol. Ecol*, **7**, 1543–1547.

Engel MS (1999) The taxonomy of recent and fossil honey bees (Hymenoptera : Apidae; *Apis*). *Journal of Hymenoptera Research*, **8**, 165–196.

Franck P, Garnery L, Celebrano G, Solignac M, Cornuet JM (2000) Hybrid origins of honey bees from Italy (*Apis mellifera ligustica*) and sicily (*A. m. sicula*). *Molecular Ecology*, **9**, 907–921.

Franck P, Garnery L, Loiseau A *et al*. (2001) Genetic diversity of the honey bee in Africa: microsatellite and mitochondrial data. *Heredity*, **86**, 420–430.

Franck P, Garnery L, Solignac M, Cornuet JM (1998) The origin of west European subspecies of honeybees (*Apis mellifera*): new insights from microsatellite and mitochondrial data. *Evolution*, **52**, 1119–1134.

Felsenstein J (1993) Phylogeny Inference Package (PHYLIP). Version 3.5. University of Washington, Seattle.

Garnery L, Cornuet JM, Solignac M (1992) Evolutionary history of the honey bee *Apis mellifera* inferred from mitochondrial DNA analysis. *Molecular Ecology*, **1**, 145–154.

Garnery L, Franck P, Baudry E *et al*. (1998a) Genetic diversity of the west European honey bee (*Apis mellifera mellifera* and *A. m. iberica*). I. Mitochondrial DNA. *Genetics Selection Evolution*, **30**, S31–S47.

Garnery L, Franck P, Baudry E *et al*. (1998b) Genetic diversity of the west European honey bee (*Apis mellifera mellifera* and *A. m. iberica*). II. Microsatellite loci. *Genetics Selection Evolution*, **30**, S49–S74.

Garnery L, Mosshine EH, Oldroyd BP, Cornuet JM (1995) Mitochondrial DNA variation in Moroccan and Spanish honey bee populations. *Molecular Ecology*, **4**, 465–471.

Garnery L, Solignac M, Celebrano G, Cornuet JM (1993) A simple test using restricted PCR-amplified mitochondrial DNA to study the genetic structure of *Apis mellifera* L. *Experientia*, **49**, 1016–1021.

Garnery L, Vautrin D, Cornuet JM, Solignac M (1991) Phylogenetic relationships in the genus *Apis* inferred from mitochondrial DNA sequence data. *Apidologie*, **22**, 87-92.

Gibert L, Scott GR, Montoya P *et al*. (2013) Evidence for an African-Iberian mammal dispersal during the pre-evaporitic Messinian. *Geology*, **41**, 691-694.

Gómez A, Lunt DH (2007) Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. In: Phylogeography of Southern European Refugia (eds Weiss S, Ferrand N), pp. 155–188. Springer, The Netherlands.

González-Sampériz P, Leroy SAG, Carrión JS *et al*. (2010) Steppes, savannahs, forests and phytodiversity reservoirs during the Pleistocene in the Iberian Peninsula. *Review of Paleobotany and Palynology*, **162**, 427-457.

Han F, Wallberg A, Webster MT (2012) From where did the Western honeybee (*Apis mellifera*) originate? *Ecology and Evolution*, **2**, 1949–1957.

Hepburn HR, Radloff SE (1996) Morphometric and pheromonal analyses of *Apis mellifera* L along a transect from the Sahara to the Pyrenees. Apidologie, **27**, 35-45.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. Nature, 405, 907–913.

Kandemir I, Özkan A, Fuchs S (2011) Reevaluation of honey bee (*Apis mellifera*) microtaxonomy: a geometric morphometric approach. *Apidologie*, 4**2**, 618–627.

Langella O (2011) Populations, version 1.2.32: Population Genetic Software. Unité Mixte de Recherches de Génétique Végétale, Institut National de la Recherche Agronomique, France.

Maguns R, Szalanski L (2010) Genetic evidence for honey bees (*Apis mellifera* L.) of Middle Eastern lineage in the United States. *Sociobiology*, **55**, 285-296.

Magnus RM, Tripodi AD, Szalanski AL (2011) Mitchondrial DNA diversity of honey bees, *Apis mellifera* L. (Hymenoptera: Apidae) from queen breeders in the United States. *Journal of Apicultural Science*, **55**, 37-46.

Meixner MD, Leta MA, Koeniger N, Fuchs S (2011) The honey bees of Ethiopia represent a new subspecies of *Apis mellifera-Apis mellifera simensis* n. ssp. *Apidologie*, **42**, 425–437.

Miguel I, Baylac M, Iriondo M *et al*. (2011) Both geometric morphometric and microsatellite data consistently support the differentiation of the *Apis mellifera* M evolutionary branch. *Apidologie*, **42**, 150–161.

Miguel I, Iriondo M, Garnery L, Sheppard WS, Estonba A (2007) Gene flow within the M evolutionary lineage of *Apis mellifera*: role of the Pyrenees, isolation by distance and postglacial re-colonization routes in the western Europe. *Apidologie*, **38**, 141–155.

Molina-Venegas R, Aparicio A, Lavergne S, Arroyo J (2015) The building of a biodiversity hotspot across a land-bridge in the Mediterranean. *Proceedings of the Royal Society of London B*, **282**, 20151116.

Muñoz I, Dall´Olio R, Lodesani M, De la Rúa P (2009) Population genetic structure of Coastal Croatian honeybees. *Apidologie*, **40**, 617-626.

Muñoz I, De la Rúa (2012) Temporal analysis of genetic diversity in a honey bee mating área of an island population (La Palma, Canary islands, Spain). *Journal of Apicultural Science*, **56**, 41-49.

Muñoz I, Pinto MA, De la Rúa P (2013) Temporal changes in mitochondrial diversity highlights contrasting population events in Macaronesian honey bee. *Apidologie*, **44**, 295-305.

Nielsen D, Page RE, Jr., Crosland MW (1994) Clinal variation and selection of MDH allozymes in honey bee populations. *Experientia,* **50**, 867-871.

Ödzil F, Yildiz MA, Hall HG (2009) Molecular characterization of Turkish honey bee populations (*Apis mellifera*) inferred from mitochondrial DNA RFLP and sequence results. *Apidologie*, **40**, 570-576.

Perrier C, Strange J, Langella O, Shepppard WS, Garnery L (2003) Diversité génétique, introgressions mitochondriales et nucléaires dans une population d'abeilles des Landes de Gascogne. *Actes du BRG*, **4**, 79–100.

Pinto MA, Henriques D, Chávez-Galarza J *et al*. (2014) Genetic integrity of the Dark European honey bee (*Apis mellifera mellifera*) from protected populations: a genome-wide assessment using SNPs and mtDNA sequence data. *Journal of Apicultural Research*, **53**, 269–278.

Pinto MA, Henriques D, Neto M *et al*. (2013) Maternal diversity patterns of Ibero-Atlantic populations reveal further complexity of Iberian honey bees. *Apidologie*, **44**, 430–439.

Pinto MA, Muñoz I, Chávez-Galarza J, De la Rúa P (2012) The Atlantic side of the Iberian Peninsula: a hot-spot of novel African honey bee maternal diversity. *Apidologie*, **43**, 663–673.

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics*, **28**, 2537–2539.

R Development Core Team (2013) R: A language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rortais A, Arnold G, Alburaki M, Legout H, Garnery L (2011) Review of the DraI COI-COII test for the conservation of the black honeybee (*Apis mellifera mellifera*). *Conservation Genetics Resources*, **3**, 383–391.

Ruttner F (1988) Biogeography and Taxonomy of Honey Bees. Springer, Berlin.

Ruttner F, Tassencourt L, Louveaux J (1978) Biometrical-statistical analysis of the geographic variability of *Apis mellifera* L. *Apidologie*, **9**, 363–381.

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular Cloning, 2nd edn. Cold Spring Harbor Laboratory Press, New York.

Sheppard WS, Meixner MD (2003) *Apis mellifera pomonella*, a new honey bee subspecies from Central Asia. *Apidologie*, **34**, 367–375.

Smith DR, Glenn TC (1995) Allozyme polymorphisms in Spanish honeybees (*Apis mellifera iberica*). *Journal of Heredity*, **86**, 12–16.

Smith DR, Palopoli MF, Taylor BR *et al.* (1991) Geographical overlap of two mitochondrial genomes in Spanish honeybees (*Apis mellifera iberica*). *Journal of Heredity*, **82**, 96–100.

Solorzano CD, Szalanski AL, Kence M *et al.* (2009) Phylogeography and population genetics of honey bees (*Apis mellifera*) from Turkey based on COI-COII sequence data. *Sociobiology*, **53**, 237-246.

Tamura K, Stecher G, Peterson D *et al.* (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*, **30**, 2725–2729.

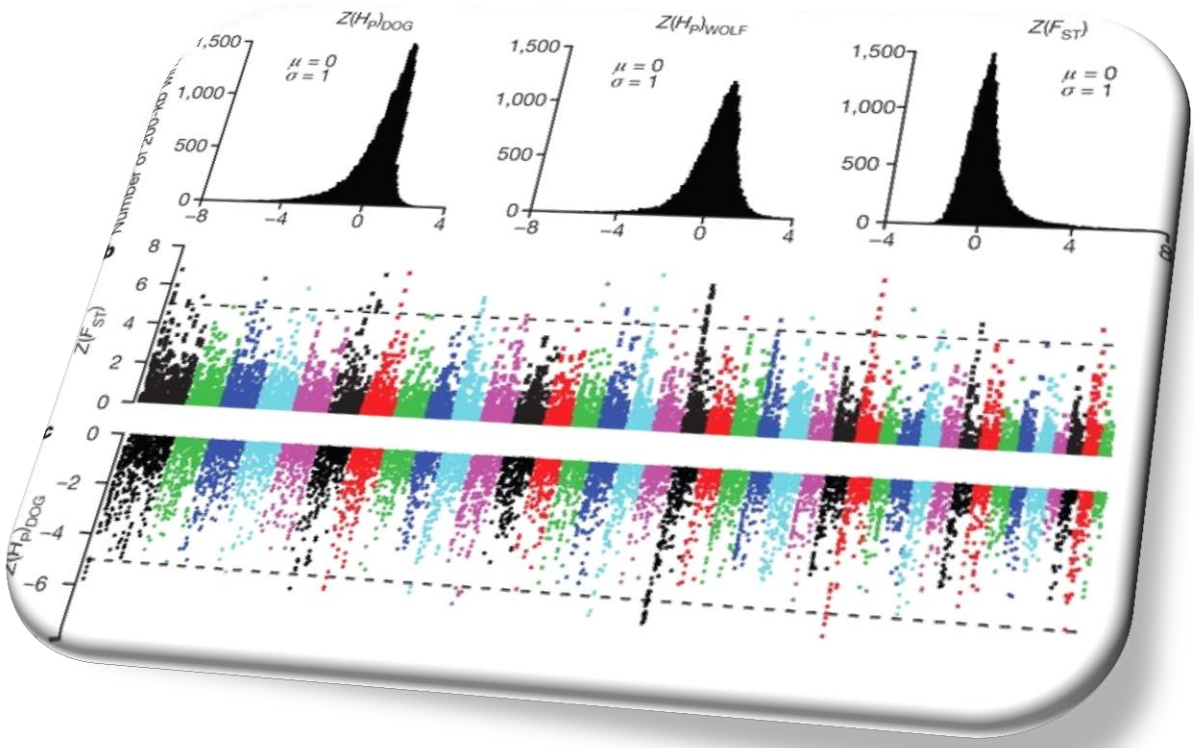
Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453-464.

Techer MA, Cleméncet J, Turpin P *et al.* (2015) Genetic characterization of the honeybee (*Apis mellifera*) population of Rodrigues Island, based on microsatellite and mitochondrial DNA. *Apidologie*, **46**, 445-454.

Wallberg A, Han F, Wellhagen G *et al*. (2014) A worldwide survey of genome sequence variation provides insight into the evolutionary history of the honeybee *Apis mellifera*. *Nature Genetics*, **46**, 1081–1088.

Weiss S, Ferrand N (2007) Current perspectives in phylogeography and the significance of South European refugia in the creation and maintenance of European biodiversity. In: Phylogeographyof Southern European Refugia (eds Weiss S, Ferrand N), pp. 237–255. Springer, The Netherlands.

Whitfield CW, Behura SK, Berlocher SH *et al*. (2006) Thrice out of Africa: ancient and recent expansions of the honey bee, *Apis mellifera*. *Science*, **314**, 642–645.



Chapter V

Signatures of selection in the Iberian honey bee (*Apis mellifera iberiensis*) revealed by a genome scan analysis of single nucleotide polymorphisms (SNPs)

The work presented in this chapter has been published:

Chávez-Galarza J, Henriques D, Johnston JS, Azevedo JC, Patton JC, Muñoz I, De la Rúa P, Pinto MA (2013) Signatures of selection in the Iberian honey bee (*Apis mellifera iberiensis*) revealed by a genome scan analysis of single nucleotide polymorphisms (SNPs). **Molecular Ecology**, 22: 5890-5907

Abstract

Understanding the genetic mechanisms of adaptive population divergence is one of the most fundamental endeavours in evolutionary biology and is becoming increasingly important as it will allow predictions about how organisms will respond to global environmental crisis. This is particularly important for the honey bee, a species of unquestionable ecological and economical importance that has been exposed to increasing human-mediated selection pressures. Here, we conducted a single nucleotide polymorphism (SNP)-based genome scan in honey bees collected across an environmental gradient in Iberia and used four F_{ST}-based outlier tests to identify genomic regions exhibiting signatures of selection. Additionally, we analysed associations between genetic and environmental data for the identification of factors that might be correlated or act as selective pressures. With these approaches, 4.4% (17 of 383) of outlier loci were cross-validated by four F_{ST}-based methods, and 8.9% (34 of 383) were cross-validated by at least three methods. Of the 34 outliers, 15 were found to be strongly associated with one or more environmental variables. Further support for selection, provided by functional genomic information, was particularly compelling for SNP outliers mapped to different genes putatively involved in the same function such as vision, xenobiotic detoxification and innate immune response. This study enabled a more rigorous consideration of selection as the underlying cause of diversity patterns in Iberian honey bees, representing an important first step towards the identification of polymorphisms implicated in local adaptation and possibly in response to recent human-mediated environmental changes.

Keywords: *Apis mellifera iberiensis*, balancing selection, directional selection, F_{ST} outlier tests, genome scan, single nucleotide polymorphism

Introduction

Patterns of genetic diversity are shaped by genome-wide (e.g. random genetic drift, gene flow) and locus-specific (e.g. natural selection) processes. While the former details population demography and phylogenetic history, the latter helps identify genes that are important for fitness and adaptation (Luikart *et al*. 2003). Accordingly, disentangling genome-wide (neutral) from locus-specific (selected) variation is a fundamental goal in evolutionary biology because the outcome not only leads to more robust inferences of demographic history but also to identification of ecologically relevant genetic variation involved in local adaptation. In the context of rapid human-induced environmental change (e.g. habitat fragmentation, climate change, introduction of novel parasites and diseases), such knowledge is becoming increasingly important for better managing and preserving genetic diversity (Allendorf & Luikart 2007).

Recent advances in computer technology and statistical genetic methods provide the tools for addressing that goal (reviewed by Luikart *et al*. 2003; Nielsen 2005; Storz 2005; Vasemägi & Primmer 2005; Helyar *et al*. 2011). Inspired by the original idea of Lewontin & Krakauer (1973) a family of statistical methods, known as the F_{ST}-based outlier tests, identify loci that exhibit frequencies significantly different than expected under neutrality for a given demographic model in genome scan data (Beaumont & Nichols 1996; Vitalis *et al*. 2001; Schlötterer 2002; Beaumont & Balding 2004; Foll & Gaggiotti 2008; Excoffier *et al*. 2009), thereby allowing separation of neutral from selected variation. The basic rationale is that genetic differentiation is higher for loci affected by directional selection (directional outliers) and lower for loci under balancing selection (balancing outliers) as compared to neutral variation. Those loci are usually called directional outliers and balancing outliers, respectively. Selection can be further validated as the cause of outlier behaviour by combining genetic and nongenetic data for the identification of environmental factors that might act as selective pressures (Foll & Gaggiotti 2006; Joost *et al*. 2007; Nielsen *et al*. 2009; Coop *et al*. 2010; Gomez-Uchida *et al*. 2011; Nunes *et al*. 2011; Prunier *et al*. 2011; Shimada *et al*. 2011; Tsumura *et al*. 2012) and by illuminating functional roles of selected loci (Luikart *et al*. 2003; Vasemägi & Primmer 2005; Prunier *et al*. 2011; Shimada *et al*. 2011; Lehtonen *et al*. 2012; Tsumura *et al*. 2012), a task that is facilitated for organisms with annotated genomes.

Owing to increasingly affordable and quicker genotyping of numerous loci scattered in the genome of numerous individuals, it is becoming increasingly popular to implement outlier tests at the genome and population-wide scales. This approach has revealed signatures of selection and, whenever supported by environmental and genomic information, has provided unprecedented

insights into the ecological and molecular basis of adaptation in numerous taxa including plants (Prunier *et al*. 2011; Tsumura *et al*. 2012), fishes (Renaut *et al*. 2011; Shimada *et al*. 2011), birds (Lehtonen *et al*. 2012), pigs (Ai *et al*. 2013) and insects (De Jong *et al*. 2013), among others. Adding to the list, in the present study, a genome-wide scan using single nucleotide polymorphisms (SNPs) was conducted in the honey bee subspecies that is native to the Iberian Peninsula: the Iberian honey bee (*Apis mellifera iberiensis*).

Apis mellifera iberiensis is one of 30 currently recognized subspecies of honey bees, which occur naturally in the Middle East, Africa and Europe (Ruttner 1988; Engel 1999; Sheppard & Meixner 2003; Meixner *et al*. 2011). Contrary to the African continent, where most honey bee subspecies are predominantly wild (Dietemann *et al*. 2009), in Europe, following arrival of the invasive mite Varroa destructor in early 1980s, wild colonies have virtually disappeared (Moritz *et al*. 2007; Jaffe *et al*. 2010). Native European honey bee subspecies (including the Iberian honey bee) are now mostly confined to apiaries.

The Iberian honey bee has been the subject of numerous population genetic surveys and thus represents one of the best studied, and yet controversial, among all subspecies. Maternal and biparental genetic markers have revealed highly complex and incongruent patterns of variation (Smith *et al*. 1991; Garnery *et al*. 1995, 1998a,b; Smith & Glenn 1995; Franck *et al*. 1998; Cánovas *et al*. 2008, 2011), which have led to competing hypotheses for the origin of the Iberian honey bee.

Early phylogeographical studies of morphology (Ruttner *et al*. 1978; Cornuet & Fresnaye 1989) and the allozyme malate dehydrogenase (Nielsen *et al*. 1994; Smith & Glenn 1995; Arias *et al*. 2006) revealed the existence of a smooth gradient extending from North Africa to France with Iberian honey bees showing intermediate phenotypes. This pattern raised the hypothesis of primary intergradation and an African origin for this subspecies (Ruttner *et al*. 1978). However, mitochondrial polymorphisms showed the co-occurrence of highly divergent African-derived (lineage A) and western European-derived (lineage-M) haplotypes forming not a smooth but a steep south–north cline in the Iberian Peninsula (Garnery *et al*. 1995; Franck *et al*. 1998; Arias *et al*. 2006; Miguel *et al*. 2007; Cánovas *et al*. 2008; Pinto *et al*. 2013), a pattern that was more consistent with a secondary contact scenario (Smith *et al*. 1991). Adding to the complexity, microsatellites (Franck *et al*. 1998; Garnery *et al*. 1998b; Cánovas et al. 2011; Miguel *et al*. 2011) and recent geometric morphometric data (Miguel *et al*. 2011) exhibited virtually no differentiation and no traces of African genes in Iberian honey bee populations and revealed a sharp break between

Iberian and northern African populations, thereby supporting neither hypothesis. The difference between maternal and bi-parental variation led Franck *et al*. (1998) to reject the secondary contact hypothesis and propose historical human-assisted introductions of African colonies with selection the best explanation for the reported morphological and allozymic clines and the diffusion and maintenance of African haplotypes in the south-western half of the Iberian Peninsula. A recent study using SNPs suggests, however, that while selection may have shaped the genome of lineage-M honey bees, which includes *A. m. iberiensis* and *Apis mellifera mellifera*, the process occurred during ancient expansions from Africa into Western Europe, resurrecting the primary intergradation hypothesis (Zayed & Whitfield 2008).

Selection has repeatedly been invoked to explain cytonuclear latitudinal patterns (Franck *et al*. 1998; Garnery *et al*. 1998a,b), yet no previous effort has attempted to evaluate its relative importance in structuring Iberian honey bee populations. In this study, populations sampled across three north–south Iberian transects were subjected to SNP-based genome scans to evaluate the importance of selection in shaping diversity patterns of Iberian honey bees. The ecological and molecular bases of the genomic regions exhibiting signatures of selection were further supported and investigated by spatial analysis, through nonrandom associations between locus-specific variation and environmental variables, and by functional annotations of genes marked by outlier SNPs. The approach followed here revealed unprecedented insights into Iberian honey bee diversity patterns, which might be helpful for future management and conservation of honey bees, an increasingly important endeavour given the current worldwide concern regarding honey bee health.

Methods

Sampling

A total of 711 honey bee haploid males, representing 23 sites (Fig. V - 1) and 237 apiaries, were collected in 2010 across three north–south transects in the Iberian Peninsula. The sites were selected to represent both the natural distribution of *Apis mellifera iberiensis* and a wide variety of climates ranging from the semi-arid in southeastern to oceanic in north-western Iberia. One transect extended along the Atlantic coast (AT, 8 sites), one through the centre (CT, 9 sites) and another along the Mediterranean coast (MT, 6 sites). The number of apiaries sampled per site varied between 8 and 13, with most sites having 10. Accordingly, sample size per site varied between 24 (8 apiaries per site x 3 hives per apiary) and 39, with most sites having 30 individuals. In each apiary, samples were taken from the inner part of three different hives and placed into

absolute ethanol. Samples were stored at -20 °C until molecular analysis. Global positioning system (GPS) coordinates were recorded in the field for each apiary.

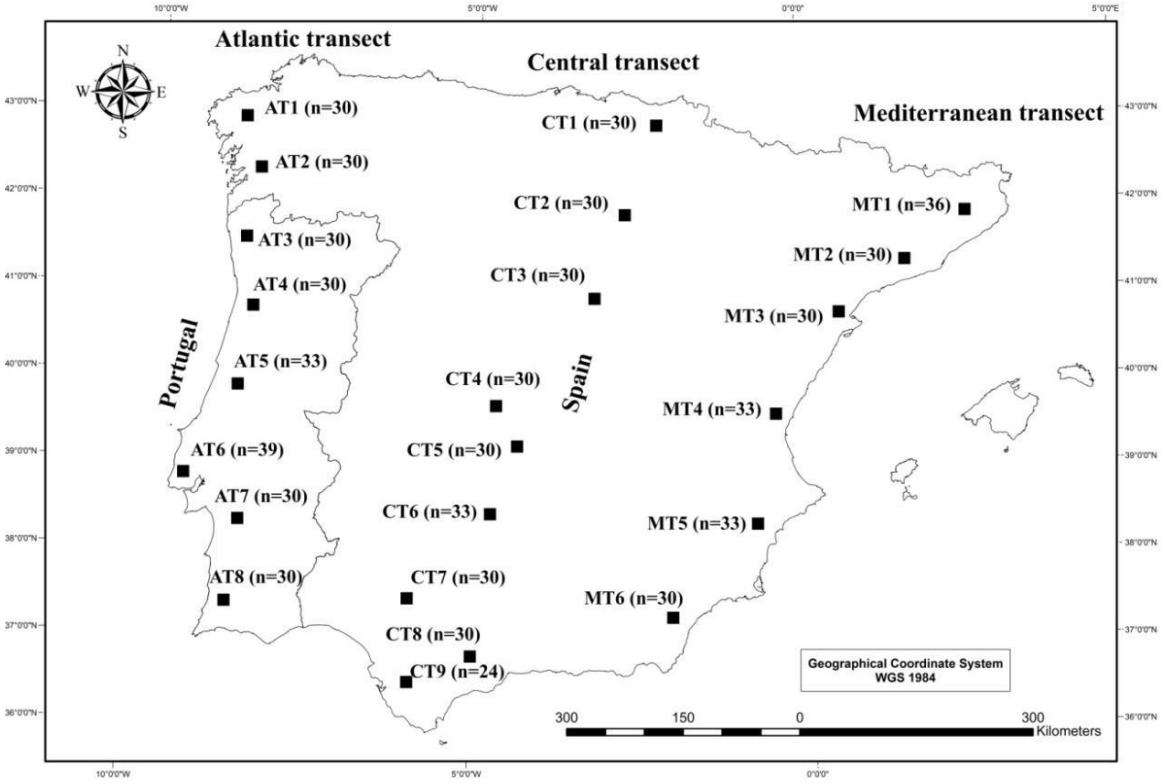


Figure V - 1. Map of the Iberian Peninsula showing the centroids of the sampling sites, sample size per site, and site codes. The number of apiaries sampled per site varied between 8 and 13, with most sites having 10.

DNA extraction and SNP genotyping

Total DNA was extracted using a phenol/chloroform isoamyl alcohol (25:24:1) protocol (Sambrook *et al*. 1989) from the thorax of 711 individuals, each representing a single colony. A total of 1536 SNP loci were genotyped for those individuals using Illumina's BeadArray Technology and the Illumina GoldenGate® Assay with a custom Oligo Pool Assay (Illumina, San Diego, CA, USA) following manufacturer's protocols.

The Oligo Pool consisted of the 768 most informative SNPs used previously to study honey bee population structure and evolution (Whitfield et al. 2006) combined with a newly developed set of 768 SNPs. Both sets of 768 SNPs were drawn from over 1.1 million SNPs defined by single base differences between (i) the reference genome of *A. mellifera* (Assembly 3.0; sequenced from the North American DH4 strain, which was primarily *A. m. ligustica*) and genome sequence traces of Africanized honey bees (largely *A. m. scutellata* admixed with the genomes of both western and

eastern European honey bees) and (ii) observed polymorphisms in ESTs. In the first case, SNPs were named 'AMB' and 'ahb', whereas in the second, they were named 'est' (Table V - 1 and Table Sup V - 1, Supporting information). The new 768 SNPs were selected with the goal of obtaining markers that were evenly spaced across the honey bee genome (D. Weaver, personal communication).

Genotype calling was performed using Illumina's GenomeStudio® Data Analysis software. For each sample, intensity clusters generated automatically by the software were manually verified and edited when necessary. SNPs with poorly separated clusters or low signals (110) were excluded from the data set. For the remaining 1426 SNPs, most honey bee samples (695 of 711) exhibited a call rate between 95% and 100%. The rest of the samples (16) had a call rate lower than 95% but above 90%.

Environmental data

Publicly available environmental data were obtained for the location of each apiary. Altitude was estimated using 30 arc-second (~1 km) spatial resolution data from the WorldClim database (http://www.worldclim.org). Climatic data were extracted from two data sets. The first data set (spatial resolution of 0.5°, representing ~50 km), which covered the period 1901–2009, was obtained from the Climatic Research Unit (www.cru.uea.ac.uk), Norwich, UK, and consisted of precipitation (Prec), minimum temperature (Tmin), mean temperature (Tmean), maximum temperature (Tmax) and cloud cover (Cld). The second data set (spatial resolution of 1°, representing ~100 km), which covered the period 1983–2005 (http://eosweb.larc.nasa.gov), was downloaded from OPENEI (http://en.openei.org), and consisted of relative humidity (Rh) and insolation (Ins), which is the amount of radiation reaching the Earth's surface per day (insolation on horizontal surface in $kWh/m^2/day$). All climatic data were integrated into a geographic information system (ArcGIS 9.3 from ESRI) to extract yearly, seasonal and monthly data. Land use/land cover data for the Iberian Peninsula was extracted from the CORINE Land Cover 2006 vector data from the European Environment Agency (http://www.eea.europa.eu). Land cover was described for each apiary by calculating the percentage of level 3 land cover classes (Heymann *et al.* 1994) within a 3 km radius circular area (28.3 km^2). To remove redundant environmental variables, that is, variables that were correlated at $|r| > 0.8$ (Manel *et al.* 2010), a principal component analysis was performed using the ade4 package (Thioulouse *et al.* 1997). Using this procedure, we kept 80

environmental variables (Table Sup V - 2, Supporting information) for further analysis, from an initial data set of 123.

Detection of outlier loci by F_{ST}-based methods

Outliers were detected using four multiple-population F_{ST}-based methods. These methods assume varying demographic models to identify loci under selection as outliers in the extreme tails of theoretical null distributions of F_{ST}. The purpose of employing conceptually different approaches was to identify potential false positives. The first method was FDIST2 (Beaumont & Nichols 1996) as implemented in LOSITAN (Antão *et al*. 2008). This coalescence-based method uses an island model to identify as outlier loci those that present unusually low or high F_{ST} values compared with neutral expectations. Approximation to the mean neutral F_{ST} in the data set was accomplished by choosing the neutral mean F_{ST} option (99% confidence interval) and running 1 000 000 simulations in LOSITAN. The second method, implemented in Arlequin 3.5.1.3, is a modification of FDIST2 that overcomes possible false positives in the presence of strong population structure by using a hierarchical island model (Excoffier *et al*. 2009). The hierarchical outlier analysis was performed by pooling the sites into an Atlantic group (all AT sites), an M-lineage group (CT1–CT3, MT1–MT3) and an A-lineage group (CT4–CT9, MT4–MT6), based on maternal patterns (Fig. Sup V - 1, Supporting information) and analysis of molecular variance (Table V - 2, Fig. Sup V - 2, Supporting information). The parameters of the run included the presence of 50 groups of 100 demes with 100 000 iterations simulated. The third method, implemented in BAYESFST, uses Markov Chain Monte Carlo simulations to assess the significance of a locus-specific parameter that indicates selection in a model of F_{ST} (Beaumont & Balding 2004). This Bayesian method was performed using 3 000 000 iterations. The fourth method, implemented in BAYESCAN 2.01, uses a Bayesian approach and a reversible-jump Markov Chain Monte Carlo method to estimate the posterior probability that a given locus is under selection (Foll & Gaggiotti 2008). It tests two alternative models: one that includes the effect of selection and another that excludes it. The BAYESCAN analysis was conducted using 20 pilot runs of 100 000 iterations, 1 500 000 iterations (sample size of 75 000 and thinning interval of 20) and an additional burn-in of 500 000 iterations.

Table V - 1. Outlier SNP loci as detected by four and three F_{ST}-based methods (BAYESFST, ARLEQUIN, and LOSITAN with *P*-values ≤ 0.05; BAYESCAN with posterior probability ≥ 0.95) and that are associated with environmental variables identified by the spatial analysis method (matSAM; CI ≥ 99%). SNP loci exhibiting the strongest signal (*P*-values ≤ 0.005 and posterior probability ≥ 0.99) are marked in bold. Genomic information for the SNP loci listed was obtained from NCBI (http://www.ncbi.nlm.nih.gov/), BEEBASE (http://hymenopteragenome.org/beebase), and FLYBASE (www.flybase.org).

SNP code	Linkage group	Position	Gene product	Putative function	SNP location	matSAM
Loci detected by BAYESFST, ARLEQUIN, LOSITAN, and BAYESCAN						
Directional						
ahb1245	1	25131116	Gst-mic2 - microsomal glutathione S-transferase 2	Metabolism	Intron	Long, Prec, Tmin, Ins
est5302	6	3199025	UDP-glucosyltransferase (UDP-glucosyltransferase 35b, Ugt35b)*	Metabolism	Exon	Lat, Prec, Tmean, Tmax, Cld
est5112	6	6341643	Vha16 - Vacuolar H+ ATP synthase 16 kDa proteolipid subunit (Vacuolar H·ATPase subunit 16-1, Vha16-1)*	Transport	3'-UTR	
ahb8266	6	13484985	Teneurin 3 – like isoform 1 (Tenascin major, Ten-m)*	Structural	Intron	Lat, Prec, Tmean, Tmax, Cld, Ins
ahb10181†	9	7162173	Hypothetical protein LOC726750/Hydrocephalus- inducing protein-like	Unknown/Structural	4110/1297	
ahb2123	10	2249429	Choline transporter-like protein 1-like	Transport	3'-UTR	Long, Lat, Prec, Ins
est7297	10	3151832	15-hydroxyprostaglandin dehydrogenase [NAD+]-like (Photoreceptor dehydrogenase, Pdh)*	Metabolism	Intron	Long, Lat, Prec, Ins
ahb2105	10	6427742	Hypothetical protein LOC100577401	Unknown	Exon	Long, Prec, Rh
AMB-00644533	11	8036193	bs-Serum response factor homolog (blistered, bs)*	Regulation	Intron	Long, Prec, Ins
est9898	13	1328979	GTP-binding protein CG1354 isoform 1	Signaling	Exon	Lat, Prec, Tmean, Tmax, Ins
est10016	13	9554011	Cytochrome P450-CYP6AS7 (Cyp6a14)*	Metabolism	Exon	Lat, Prec, Tmean, Tmax, Cld, Ins
est11018	15	5299405	NimC2-nimrod C2 (nimrod C2, nimC2)*	Immunity	5'-UTR	
Balancing						
ahb142	1	21550552	Sema 1-Semaphorin 1A (Sema-1a)*	Structural	Intron	
ahb6903	4	9847583	Dscam-Down syndrome cell adhesion molecule (Down syndrome cell adhesion molecule, Dscam)*	Immunity	Intron	
AMB-00963630†	12	1214828	Hypothetical protein LOC100576488/Collagen alpha-2(IX) chain-like	Unknown/Structural	18439/25703	
AMB-00708602	15	8856798	Cubilin-like	Signaling	Intron	
AMB-00914134‡	0	78294	Protein lin10-like (X11Lβ)*	Regulation	6219	

Loci detected by BAYESFST, ARLEQUIN, and LOSITAN						
Directional						
AMB-00905664	1	15234278	Hypothetical protein LOC100578389	Unknown	Intron	
ahb1232[†]	1	25169032	Notum pectinacetylesterase homolog (Notum)*/ NMDA kainate 2 sensitive receptor	Signaling/Signaling	21596/10461	
ahb226[§]	1	29654719				
est2423	2	2983426	Retinol dehydrogenase 11-like	Metabolism	Exon	Ins
AMB-00190928	2	7469407	Ubc-E2H-Ubiquitin-conjugating enzyme E2 H (UbcE2H)*	Signaling	Intron	Lat, Prec, Ins
AMB-00402575	3	5547433	Protein outspread-like (outspread, osp)*	Regulation	Intron	
est5553	7	4067968	Aldh-Aldehyde dehydrogenase isoform 1 (Aldehyde dehydrogenase, Aldh)*	Metabolism	5'-UTR	Long
est6087	8	5157656	Rfabg-Retinoid and fatty acid-binding glycoprotein isoform1 (Retinoid- and fatty acid-binding glycoprotein, Rfabg)*	Transport	Exon	
ahb9731	8	8266264	Hypothetical protein LOC411273	Unknown	Exon	Lat, Prec, Ins
ahb10154	9	7170480	Hydrocephalus- inducing protein-like	Structural	Exon	Lat, Prec, Tmean, Ins
AMB-00119087[§]	0	30896				Long, Prec
Balancing						
ahb1129	1	18156863	Hypothetical protein LOC413562	Regulation	Intron	
est5796	7	5788241	Hypothetical LOC100578906	Unknown	5'-UTR	
est6265	8	6383127	PHD finger and CXXC domain-containing protein CG17446-like isoform1 (Cfp1)*	Regulation	3'-UTR	
AMB-00338679	8	11725179	Midasin-like	Structural	Intron	
AMB-00310216[‡]	10	5583063	5 HT2 beta-Serotonin receptor	Signaling	22473	
ahb4188	14	9203355	Blop-blue-sensitive opsin (Rhodopsin, Rh5)*	Signaling	Exon	

Lat, latitude; Long, longitude; Prec, precipitation; Tmin, minimum temperature; Tmean, mean temperature; Tmax, maximum temperature; Ins, insolation; Cld, cloud cover; Rh, relative humidity.

*Names and/or symbols within parentheses correspond to orthologous genes of Drosophila melanogaster as in FLYBASE.

[†]SNP located between two genes (or putative genes). Function of both genes and physical distances (bp) to the 3′ or 5′ ends are indicated.

[‡]SNP located close to a gene. Physical distance (bp) to the 3′ or 5′ end of the gene is indicated in the column 'SNP location'.

[§]SNP located far away from genes (>167.8 kb).

Introgression and ascertainment bias

Because introgression may mimic selection, prior to the F_{ST}-based tests the 711 Iberian individuals were assessed for introgression using Structure 2.3.3 (Pritchard *et al*. 2000) and by implementing a principal components analysis with ADEGENET 1.3–7 (Jombart 2008). Over 1075 polymorphic SNPs and a reference collection consisting of the two eastern European beekeepers-favourite honey bee subspecies, *Apis mellifera ligustica* (17 individuals) and *Apis mellifera carnica* (19 individuals) and the northern African subspecies *Apis mellifera intermissa* (31 individuals) were used. These analyses revealed virtually no signs of contemporaneous or historical introgression from the two eastern European or the northern African subspecies (Figs. Sup V -3 and Sup V - 4, Supporting information for the parameters settings and results), suggesting that introgression should not be a confounding factor in the F_{ST}-based outlier detection. Retaining uninformative monomorphic loci in the data set tends to increase dramatically the rate of false positives when searching for selection using F_{ST}-based methods (Nielsen *et al*. 2009; Gomez-Uchida *et al*. 2011). Accordingly, the Iberian data set was screened for monomorphic SNPs, as defined by a cut-off criterion of >0.98 for the most common allele. This filtering process produced a final data set of 383 polymorphic loci (288 genomic and 95 EST-derived) for the Iberian honey bee, which was used in all subsequent analyses performed in this study. The high proportion of monomorphic SNPs in Iberian honey bees might be explained by ascertainment bias. While the SNP panel was relatively diverse it did not include *A. m. iberiensis*. Accordingly, when SNPs were genotyped on Iberian honey bees, an ascertainment bias was introduced. However, ascertainment bias is expected to affect every genotyped Iberian individual equally and thus not systematically bias any particular individual or population, and thus evidence for selection is not caused by a SNP discovery artefact.

Genomic information

Each outlier SNP's 100 bp flanking sequence was mapped to the Honey Bee Assembly 4.5 using BLAST in BEEBASE (hymenopteragenome.org/beebase) and NCBI (www.ncbi.nlm.nih.gov). Genomic position was ascertained using the Map Viewer tool available in NCBI. SNPs were classified as belonging to exons, introns, 3′ or 5′ untranslated region (UTR), or intergenic regions. Genes marked by SNPs were identified using the Official Gene Set 3.2 (BEEBASE) and Entrez Gene (NCBI). As functional annotation of the honey bee genome is incomplete, putative Gene Ontology classifications were ascribed to as many genes as possible, based on homology to *Drosophila*

melanogaster, using best-BLASTP hit and e-value cut-off 0.01 in FLYBASE (www.flybase.org) complemented by NCBI annotation.

Differentiation of neutral and directional outlier loci

Differentiation of putatively neutral and the strongest directional outliers was investigated by analysis of molecular variance (AMOVA). Two- and three-level AMOVAs were performed using Arlequin (Excoffier & Lischer 2010) with 10 000 permutations to assess whether levels of differentiation were significantly >0. Two-level AMOVAs (a single group) were conducted to test the prediction of a higher variance attributed to sites for directional outliers as compared to putatively neutral loci. Three-level AMOVAs (multiple groups) were conducted to assess (i) whether putatively neutral and outlier variation were similarly structured and, if there were any structure, (ii) whether neutral and/or outlier structure were oriented predominantly east–west or north–south. To that end, each neutral and directional outlier data set was partitioned by transect (longitudinal grouping) and mtDNA lineage (latitudinal grouping) into different grouping combinations (Fig. Sup V - 2, Supporting information for details) in an effort to search for structures that generated the highest variance fraction attributed to groups.

Associations between environmental variables and outlier loci

A spatial analysis was performed to identify associations between allelic frequencies and environmental variables by using the software matSAM (Joost *et al*. 2008). This approach may provide insightful clues about selective forces acting upon outlier loci. Multiple univariate logistic regression models were computed by matSAM at the individual level (representing a single hive). The software matSAM assesses the significance of the coefficients calculated by the logistic regression function by implementing likelihood ratio (G) and Wald statistical tests. A model is considered significant only if the null hypothesis is rejected by both tests, after Bonferroni correction (Joost *et al*. 2007). The 383 SNP data set was tested against 80 environmental variables described previously (see Table Sup V - 2, Supporting information). The significance threshold level was set to 1.632E-7, corresponding to a 99% confidence interval following Bonferroni correction. In addition to the matSAM analysis performed at the apiary level, associations between land cover and the strongest directional outlier loci were further examined for the 23 sites through linear regressions of allelic frequencies and percentage of level 3 land cover classes within a circular area of 1963.5 km^2 (25 km radius) around each site's centroid. In this

analysis, different combinations of agriculture land cover classes were tested, and only the best models were selected.

Table V - 2. Variance components (%) of the three-level AMOVAs obtained with the 309 neutral loci and the nine strongest directional outliers for different grouping combinations (all *P-values* < 0.0001). Grouping by transects and mtDNA lineage were the main criteria (see Fig. Sup V - 2 for a graphical visualization of the five groupings).

Grouping criteria	N° of Groups	Within sites		Among sites		Among groups	
		Neutral	Outlier	Neutral	Outlier	Neutral	Outlier
By mtDNA lineage[a]	2	95.75	80.14	3.35	11.56	0.89	8.29
By transect[b]	3	96.01	81.24	3.24	8.05	0.75	10.71
By AT & mtDNA lineage[c]	3	95.92	80.70	3.05	6.59	1.03	12.70
By AT & mtDNA lineage[d]	3	95.95	81.32	3.09	8.17	0.96	10.51
By transect & mtDNA lineage[e]	6	96.05	82.19	2.89	5.53	1.06	12.28

[a] Group 1: all M-lineage sites across transects; Group 2: all A-lineage sites across transects
[b] Group 1: all AT sites; Group 2: all CT sites; Group 3: all MT sites
[c] Group 1: all AT sites; Group 2: M-lineage sites of CT and MT; Group 3: A-lineage sites of CT and MT
[d] Group 1: only A-lineage AT sites; Group 2: M-lineage sites of AT, CT and MT; Group 3: A-lineage sites of CT and MT
[e] Group 1: M-lineage site of AT; Group 2: A-lineage sites of AT; Group 3: M-lineage sites of CT; Group 4: A-lineage sites of CT; Group 5: M-lineage sites of MT; Group 6: A-lineage sites of MT

Results

Detection of outlier loci by F_{ST}-based methods

The genome scan approach implemented in this study identified, by at least one of the four F_{ST}-based methods, a total of 69 outlier loci (of 383; 18.0%) at a 95% confidence level (Table V - 1 and Table Sup V - 1, Supporting information). The detection rate of outliers varied among the four methods, with the highest number of loci obtained by LOSITAN (57 of 383; 14.9%), and the lowest by BAYESCAN (17 of 383; 4.4%). The approaches implemented by Arlequin and BAYESFST detected 49 (12.8%) and 41 loci (10.7%), respectively. All loci detected by BAYESCAN were also identified by the other three methods, whereas 15 loci were exclusive to LOSITAN, six to Arlequin, and four to BAYESFST. As predicted from other studies (Wilding *et al*. 2001; Bonin *et al*. 2006; Shimada *et al*. 2011; Wang *et al*. 2012), the number of outlier loci exhibiting a signal of directional selection (50) was higher than that of balancing selection (19; Table V - 1 and Table Sup V - 1, Supporting information) and varied according to the F_{ST}-based method employed. Specifically, the number of outlier loci under directional and balancing selection was as follows, respectively: LOSITAN 43/14, Arlequin 33/16, BAYESFST 27/14 and BAYESCAN 12/5 (Table V - 1, Table Sup V - 1 and Fig. Sup V - 5, Supporting information).

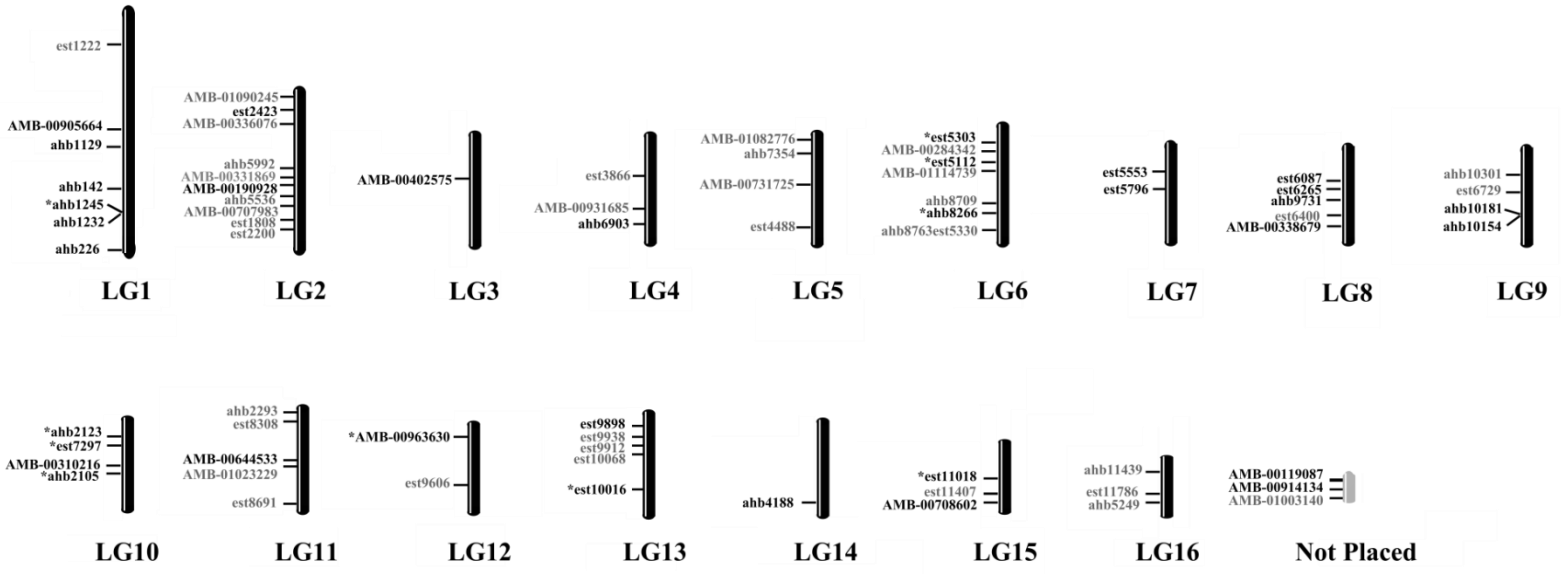


Figure V - 2. Physical map of the 16 honey bee linkage groups (LG1 to LG16) showing the genomic positions of the 69 outlier SNP loci detected by at least one F_{ST}-based method and that are associated with at least one environmental variable. The 34 outliers detected by at least three F_{ST}-based methods are marked in bold (Table V - 1) whereas those detected by two or fewer F_{ST}-based methods (Table Sup V - 1, Supporting information) are marked in gray. The 10 strongest candidates for selection (Table V - 1) are marked with an asterisk. Three SNPs have no position assigned yet. The map was depicted from the honey bee genome sequence available at http://www.ncbi.nlm.nih.gov/projects/mapview using the Map Viewer tool.

The 69 outliers were dispersed throughout the 16 honey bee linkage groups (LG) with three loci still unplaced in a chromosome (Fig. V - 2). LG 1, 2 and 6 harboured the highest number of outlier loci with 7, 10 and 7, respectively. The lowest physical distance (linkage) between outlier loci was observed for pairs ahb1245/ahb1232 (LG1) and ahb10181/ahb10154 (LG9) with 37.9 and 8.3 kb, respectively. The remaining 65 loci were further apart, with physical distances varying between 65.9 and 26 304 kb for pairs est9938/est9912 (LG13) and est1222/ahb226 (LG1), respectively. Therefore, except for the two pairs above, given the honey bee's exceptionally high recombination rate of 23.2 cM/ Mb (Beye *et al*. 2006) most outlier loci might be considered unlinked. Indeed, in an earlier study using 1136 SNPs (many employed herein) across 14 honey bee subspecies and Africanized honey bees, Whitfield *et al*. (2006) observed a rapid decay of linkage disequilibrium over a distance of 5–10 kb.

Among the 69 outlier loci, 34 (23 directional and 11 balancing) were detected by at least three F_{ST}-based methods (Table V - 1), whereas 35 were detected by two or fewer F_{ST}-based methods (Table Sup V - 1, Supporting information) and might be false positives. In contrast, 17 loci are strong candidates for selection as they were simultaneously detected by the four methods (Table V - 1). Among these 17, 10 loci (nine directional and one balancing) exhibited the strongest signal (P-value ≤ 0.005 and posterior P ≥ 0.99) being therefore the best candidates (marked in bold in Table V - 1). The nine best directional outliers exhibited large differences in allele frequencies among sites, contrasting with the sole candidate (AMB-00963630) for balancing selection with nearly even frequencies across the geographical range (Fig. Sup V - 1, Supporting information). In spite of the weaker signal, 17 additional loci can still be considered good candidates as they were detected by three F_{ST}-based methods (Table V - 1).

Genomic information shows that of the 34 outliers, 32 loci were located in or near genes (Table V - 1) that code for proteins (Table Sup V - 3, Supporting information for accession numbers) involved in a diverse array of putative functions including signalling, structural, metabolism, regulation, transport and immunity. Among the 27 outliers that mapped to genes, 12 were located in introns, 6 in untranslated regions (3′ or 5′ UTR) and 9 in exons, although none was predicted to induce amino acid changes.

Differentiation of neutral and directional outlier loci

To assess whether putatively neutral and directional outlier variation were similarly structured, two- and three-level hierarchical AMOVAs were performed for different loci combinations (neutral vs. the

strongest outliers). For the two-level AMOVA, the percentage of the total variance among sites was substantially higher for outliers (15.97%; P < 0.0001) than for neutral loci (3.77%; P < 0.0001). As observed for the two-level AMOVA, when multiple groups were considered (three-level AMOVA), the variance component attributed to groups was higher for outlier loci than for neutral loci, regardless of the groupings tested (Table V - 2).

For neutral loci, the variance due to groups was low (0.75–1.06%) and always smaller than the variance among sites (2.89–3.35%). The best variance partitioning was obtained for six groups (1.06), although slightly better (1.03) than that obtained for three groups formed by the Atlantic sites, the M-lineage sites of central and Mediterranean transects and the A-lineage sites of central and Mediterranean transects (Table V - 2, Fig. Sup V - 2, Supporting information). Separating sites in the central and Mediterranean transects, using the same criteria applied to the Atlantic sites, generated poorer variance partitioning (data not shown). Overall, these results indicate a very weak neutral substructure across the Iberian Peninsula, although the Atlantic populations seem to be slightly more differentiated.

In contrast to the result with neutral loci, when the strongest directional outliers were analyzed, the variance component attributed to groups was elevated (10.51– 12.70%) and was higher than the variance among sites (5.53–8.17%), as long as the Atlantic transect was kept as a separate group. Grouping the Atlantic with central and/or Mediterranean sites always inflated the variance attributed to sites and lowered the among groups component (Table V - 2 and data not shown). In summary, these results indicate that while neutral structure is very weak, directional outlier structure is pronounced and stronger longitudinally than latitudinally.

Associations between environmental variables and outlier loci

In addition to the F_{ST}-based outlier approaches, genomic regions exhibiting a signal of selection were further confirmed by the spatial analysis implemented by matSAM at the individual level. Significant associations (≥99% confidence level with Bonferroni correction), with at least one environmental variable, were detected for 33 (8.62%) out of the 383 screened loci (Table Sup V - 2, Supporting information). Of the 33 loci, five were exclusive to matSAM and 28 had previously been identified by at least one F_{ST}-based method as directional outliers (Table V - 1 and Table Sup V - 1, Supporting information).

The environmental variables that were more frequently associated with SNPs were precipitation and longitude, with 21 and 17 such associations, respectively (Table V - 1, Tables Sup

V - 1 and Sup V - 2, Supporting information). The variables altitude and land cover were on the opposite side of the spectrum with no significant (after Bonferroni correction) associations detected. Most loci (20) were associated with a single or two variables. A few loci (four) were associated with more than four variables. The greatest number and strongest associations were found for locus ahb8266, which was correlated with latitude, precipitation, minimum temperature, mean temperature, maximum temperature, cloud cover and insolation (Table V - 1).

Table V - 3. Best linear regression models between the nine strongest directional outliers (see Table V - 1) and land cover at the site scale.

SNP code	Land cover classes[1]	r	r^2	*P*-value
ahb1245	None	-	-	-
ahb2123	None	-	-	-
est7297	None	-	-	-
ahb2105	222	0.4330	0.1875	0.03902
est11018	222	0.6415	0.4116	0.00097
est5112	211	0.5846	0.3417	0.00339
est5302	211+212+222+223+241	0.5391	0.2907	0.00794
ahb8266	211+212+222+223+241+242	0.5567	0.3099	0.00580
est10016	211+212+222+223+241+242+244	0.6591	0.4344	0.00062

[1]Corine land cover 2006 codes: (211) non-irrigated arable land, (212) permanently irrigated land, (222) fruit trees and berry plantations, (223) olive groves, (241) annual crops associated with permanent crops, (242) complex cultivation patterns, and (244) agro-forestry areas. For each site, land cover classes were considered individually and combined within a 20 km radius circle. Regression analyses tested for associations between allele frequencies and the land cover proportion for the 23 sites.

Among the 34 loci that were detected by at least three and four F_{ST}-based methods, 15 and 9 loci, respectively, showed significant (≥ 99%) associations with at least one environmental variable (Table V - 1). Overall, only nine loci were simultaneously detected by the five methods, with confidence levels above 99.99%. These loci were the strongest candidates for directional selection and were mainly associated with latitude, longitude, precipitation and insolation.

Regression analyses of allele frequencies and land cover implemented at the site level revealed significant associations for six of the nine strongest directional outliers (Table V - 3). The strongest correlations were detected for loci est10016, est11018, and est5112. Loci ahb1245, ahb2123 and est7297 were not associated with any of the land cover classes tested. The land classes that, individually or combined, produced the best models were all agricultural and included 'nonirrigated arable land', 'permanently irrigated land', 'fruit trees and berry plantations', 'olive groves', 'annual crops associated with permanent crops', 'complex cultivation patterns' and 'agro-

forestry areas'. None of the forest classes tested was found to be associated with allele frequencies (data not shown).

Discussion

Candidate SNP loci for selection

Determining whether an outlier is a marker of a selected locus or a false positive is a major concern when searching for adaptive molecular variation. A common strategy has been to seek for confirmatory evidence from multiple outlier approaches that generate the expected neutral distribution of F_{ST} estimates under different demographic scenarios (Luikart *et al*. 2003; Storz 2005; Vasemägi & Primmer 2005). In this study, of the 69 detected outliers, only 17 (all identified by BAYESCAN) were cross-validated by the four conceptually different methods, whereas 34 and 44 (Table V - 1 and Table Sup V - 1, Supporting information) were simultaneously detected by at least three and two methods, respectively. While discrepant results among outlier approaches have been repeatedly reported (Bonin *et al*. 2006; Shikano *et al*. 2010; Nunes *et al*. 2011; Shimada *et al*. 2011; De Jong *et al*. 2013), comparative simulations have shown that BAYESCAN outperforms under a wide range of scenarios, exhibiting the lowest rate of false positives (Pérez-Figueroa *et al*. 2010; Narum & Hess 2011). Accordingly, these 17 outliers are the best candidates for selection. At the same time, the proportion of outlier loci detected by at least four (4.4%), three (8.9%) or two (11.5%) methods is within the range reported for other taxa (extensively reviewed by Nosil *et al*. 2009; Shikano *et al*. 2010; Nunes *et al*. 2011; Shimada *et al*. 2011; Buckley *et al*. 2012; among others), including honey bees (Zayed & Whitfield 2008), and all are potentially important markers, including 10 loci identified by only two methods that map to functional variation, which deserve further attention to avoid the risk of losing interesting candidates.

A common feature of outlier approaches is their higher power in detecting directional selection compared with balancing selection, a disparity that is particularly pronounced when levels of genetic differentiation are low (Beaumont & Balding 2004; Foll & Gaggiotti 2008). While this would explain a higher proportion of directional than balancing outliers detected herein and in most studies (Nielsen *et al*. 2009; Cooke *et al*. 2012; Limborg *et al*. 2012), the fact that five loci were identified by every single method in a scenario of low differentiation (F_{ST} values ranged between 0.008 and 0.093 for the 383 loci; data not shown) across the Iberian honey bee range indicates a strong signal of balancing selection, especially for locus AMB-00963630, which was detected with a confidence level above 99%.

Biological relevance of candidate genes and association with possible selection pressures

Support for selection comes from functional annotations of genes carrying outlier SNPs that relate directly to colony fitness, and their association with possible selection pressures identified by the spatial analysis (Table V - 1). Support is particularly compelling for outlier loci mapped to different genes putatively involved in the same function, which is the case of balancing and directional outliers that are mapped to genes encoding proteins that are related to vision, xenobiotic detoxification and immune response.

Outlier SNPs mapped to genes related to vision. Five outlier SNPs mark genes associated with vision. The strong directional outlier ahb8266 is mapped to an intron of a gene that encodes a transmembrane glycoprotein named teneurin 3-like isoform 1 (GB12816). Its putative orthologue in *Drosophila melanogaster*, tenascin major, has been shown to accomplish an important role during neural development and in the process of vision (Kinel-Tahan *et al*. 2007). Balancing outlier ahb142 is also located in an intronic region of a transmembrane protein named Semaphorin 1A (GB11468), which is required for the synapse formation and axon guidance (Godenschwege *et al*. 2002). This protein has been shown to participate in regulating the photoreceptor axon guidance in the visual system of *D. melanogaster* establishing an appropriate topographic termination pattern in the optic lobe (Cafferty *et al*. 2006). Three additional outliers, two directional (est7297, est2423) and one balancing (ahb4188), are putatively related to vision through their participation in the visual cycle, which has been shown to be a major pathway contributing to the maintenance of rhodopsin levels in *D. melanogaster* (Wang *et al*. 2010). The strong outlier est7297 is located in the intronic region of a gene that encodes a 15-hydroxyprostaglandin dehydrogenase [NAD+]-like protein (GB11685) belonging to NADB-Rossman superfamily, based on sequence comparison. Its putative orthologue in *D. melanogaster* is a photoreceptor dehydrogenase that participates in regenerating the chromophore for the production of rhodopsin (Wang *et al*. 2010). Outlier est2423 is located in the exon of a gene that encodes a retinol dehydrogenase 11-like protein (GB11195), which might be implicated in the production of the chromophore by catalysing the cis-retinol to cis-retinal (Kim *et al*. 2005; Belyaeva *et al*. 2009; Wang et al. 2010). Finally, balancing outlier ahb4188 is mapped to the exon of a gene encoding for a blop (blue sensitive) opsin (GB13493), which accomplishes a receptor function of the chromophore and forms a rhodopsin responsible for absorbing the 440 nm (blue spectra) wavelength (Townson *et al*. 1998; Earl & Britt 2006). Fixation of an alternative allele could confer better protein activity or structural conformation (Camps *et al*.

2007) for the production and regeneration of the chromophore, while the maintenance of polymorphism could provide an adaptive advantage favouring the absorbance of different wavelengths within blue spectra, as recently reported for titi monkeys (Bunce *et al*. 2011).

The three directional outliers (ahb8266, est7297, and est2423) are associated with insolation, among other environmental variables, suggesting that these loci mark genomic regions that are involved in local vision-associated adaptation. Vision is a key component of foraging behaviour (Winston 1987). Flight departures from the colony for foraging have been both positively and negatively correlated with solar radiation intensity (Burrill & Dietz 1981). Once outside, foragers rely on their highly developed trichromatic visual system and use colour discrimination for finding food sources and homing to the hive (reviewed by Menzel & Muller 1996). It is therefore possible that selection favours alleles that enable a more efficient light perception at contrasting climates such as the Mediterranean with long periods of sunny skies in southern Iberia, and the Atlantic with long periods of rain and cloudy skies, particularly in north-western Iberia.

Outlier SNPs mapped to genes related to xenobiotic detoxification. Three directional SNPs mark genes involved with detoxification of xenobiotics. Outlier est10016 is located in the exon of a gene that encodes CYP6AS7 protein (GB18052), which belongs to the cytochrome P450s monooxygenases superfamily. This protein superfamily plays a major role in the protection against xenobiotics and has been implicated in tolerance to plant toxins (Scott & Wen 2001; Mao *et al*. 2009) and evolved resistance to pesticides in many insects (Feyereisen 1999; Ffrench-Constant *et al*. 2004; Li *et al*. 2007), including tolerance of pyrethroid insecticides in honey bees (Pilling *et al*. 1995; Johnson *et al*. 2006; Mao *et al*. 2011). Outlier ahb1245 is located in an intron of a putative gene that codes for a microsomal glutathione-S-transferase 2 (GB10566), which belongs to another major superfamily associated with detoxification, the glutathione-S-transferases (GSTs). This gene has been implicated in the detoxification of metabolites formed by cytochrome P450 enzymes (Yu 2002; Claudianos *et al*. 2006). Finally, outlier est5302 maps to the exon of a gene (GB10566) that codes for a protein belonging to the UDP-glycosyltransferases (UGT) superfamily whose probable function is related to olfaction and detoxification mechanisms, based on its putative *D. melanogaster* UDP-glucosyltransferase 35b orthologue (Bull & Whitten 1972; Wang *et al*. 1999). In addition to the xenobiotic metabolism, these genes might also play a role in the defence against pathogens. A study in *D. melanogaster* suggested that genes encoding GSTs and cytochromes P450 likely participate in the detoxification of reactive oxygen species produced during microbial

killing to protect the gut epithelium (Buchon *et al.* 2009). A similar mechanism might be involved in honey bees, as suggested by an elevated GST activity in the honey bee gut infected by one of its most important pathogens, the microsporidium *Nosema ceranae* (Dussaubat *et al.* 2012).

It makes perfect sense that three outlier SNPs map to loci involved with detoxification. During their lifetime, honey bees are exposed to a wide range of natural (e.g. phytochemicals present in nectar, pollen and propolis) and synthetic (agricultural pesticides) xenobiotics. While honey bees have evolved detoxification mechanisms to metabolize natural phytochemicals, they are known for being unusually sensitive to a range of insecticides (Hardstone & Scott 2010). Agricultural insecticides used for crop protection are increasingly recognized as major drivers of recent worldwide honey bee losses. In 2012 alone, more than 100 papers and reports were published on the impact of insecticides, particularly neonicotinoids, on bees (mentioned by Osborne 2012). In addition to acute poisoning accidents (e.g. the case in Germany described by vanEngelsdorp & Meixner 2010), neonicotinoids seem to have a pervasive impact on honey bees through short- and longterm chronical exposure to sub-lethal doses that can result in reduction of disease resistance and breeding success as well as behavioural disturbances including problems in flying and navigation, impaired memory and learning, and reduced foraging ability (extensively reviewed by Blacquière *et al.* 2012). Accordingly, honey bees are potentially under strong selection pressure when they forage in intensive farming areas. In this study, we observed significant correlations between agricultural land cover (a surrogate of agrochemicals use) and outlier SNPs est10016 and est5302 (Table V - 3). Given increasing evidence of sub-lethal effects of pesticides on colony survival, this finding deserves further investigation.

Outlier SNPs mapped to genes related to innate immune response. Two strong outlier SNPs, which showed contrasting selection signatures, mark genes implicated in immune response. While est11018 exhibits a signal of directional selection and ahb6903 is seemingly under balancing selection, both modes of selection have been shown to be associated with immune response in fruit fly (Jiggins & Hurst 2003; Sackton *et al.* 2007), cod (Gomez-Uchida *et al.* 2011) and humans (Hollox & Armour 2008). Directional outlier est11018 maps to the 5′UTR region of a gene that encodes a putative Nimrod C2 protein (GB13979). Nimrod C2 belongs to a diverse class of transmembrane proteins that have been shown to function as phagocytosis receptors and/or microbial binding factors, suggesting an important role in the cellular immunity and elimination of apoptotic cells (Kurucz *et al.* 2007; Sackton *et al.* 2007; Somogyi *et al.* 2008). Balancing outlier

ahb6903 is located in the intron of a gene encoding a Dscam-Down syndrome cell-adhesion molecule (GB15141). *DSCAM*, an orthologous gene in *D. melanogaster*, encodes for a receptor that has long been implicated in neuronal development, although more recently it has also been shown to participate in the phagocytosis of infectious non-self (Watson *et al. 2005*). As expected, the two outliers were not associated with any of the environmental variables tested in this study. Despite potential direct influence of climate on both host and pathogen, selection pressure on genomic regions related to immune response is imposed by parasites and pathogens, which in a managed population such as the honey bee, will be more dependent on beekeeping activities.

As was true with detoxification, selection for alleles associated with immunity makes biological sense. Honey bees face a multitude of enemies which, in addition to pesticides, are recognized as key contributors to worldwide population declines (reviewed by vanEngelsdorp & Meixner 2010; Potts *et al*. 2010). While honey bees have evolved both group (e.g. grooming, nest hygiene, necrophoric behaviour, behavioural fever) and individual (mechanical, physiological and immune responses) defense mechanisms, their health has been challenged in unprecedented ways by an increasing number of emerging enemies including parasites, pathogens and predators. Among the most significant are the parasitic mite *Varroa destructor*, the gram-positive bacterium *Paenibacillus larvae* and the microsporidia *Nosema apis* and *N. ceranae* (vanEngelsdorp & Meixner 2010). The latter has been particularly worrisome in Spain, where it has been reported as a highly virulent pathogen and a potential culprit in colony collapse disorder (Higes *et al*. 2009, 2010), although epidemiological studies in the US (Cox-Foster *et al*. 2007; vanEngelsdorp *et al*. 2009) and Germany (Gisder *et al*. 2010) failed to associate this emerging enemy to colony losses. While the impact of *V. destructor* (on its own and as a vector of viruses) has been alleviated by in-hive acaricides (although mite-resistance has been reported for an increasing number of chemicals), colonies are rarely treated against the pathogens *P. larvae* and *Nosema* spp. because antibiotics are interdicted in the European Union due to honey contamination. Accordingly, Iberian honey bees have likely been under pathogen-driven selective pressure, although only two strong outliers, both mapped to genes putatively involved in phagocytosis immune response, were identified in this study. In *D. melanogaster*, genes implicated in phagocytosis have been found to have an important role in deterring bacteria (Kurucz *et al*. 2007) and fungi (Jin *et al*. 2008). While little is known about the importance of phagocytosis in honey bees (Evans & Spivak 2010), gene-expression studies have reported differential regulation of *NIMC2* or *DSCAM* genes in colonies exposed to

different parasites and pathogens (Evans 2006; Navajas *et al*. 2008; Bull *et al*. 2012; Nazzi *et al*. 2012) suggesting that these genes might participate in the immune response.

The SNP outliers identified in this study are only a small fraction of the total SNP variation in the honey bee. As such, they are markers of selection, and unlikely to be direct targets of selection. Many of the outlier SNPs were located in introns and UTRs and likely have little or no impact on the polypeptide, although these same regions may be involved in regulation or alternative splicing (Ladd & Cooper 2002). Those outlier SNPs that were located in exons were all silent substitutions, which have widely been assumed to have no effect on protein fitness, although comparison of all dipteran and hymenopteran sequenced genomes, shows that codon usage bias is highest in the honey bee (Behura & Severson 2012) and is associated with differential translation rates, mRNA stability and modification of protein structure and activity (Tsai *et al*. 2008). A few SNPs were located outside coding regions. Accordingly, it is likely that many SNPs, if not all, have hitchhiked with linked polymorphisms that are the causative mutations under selection.

Differentiation of neutral and directional outlier loci

Partition of neutral variation was negligible across the Iberian honey bee range, barely reflecting the steep north–south maternal cline formed by the two highly divergent western European (M) and African (A) lineages extensively reported in the literature (Garnery *et al*. 1995; Franck *et al*. 1998; Arias *et al*. 2006; Miguel *et al*. 2007; Cánovas *et al*. 2008; Pinto *et al*. 2013) and further confirmed in this study (Fig. Sup V - 1, Supporting information). A pattern of virtually no differentiation was also revealed by geometric morphometric data (Miguel *et al*. 2011) and microsatellites (Franck *et al*. 1998; Garnery *et al*. 1998b; Cánovas *et al*. 2011) but not by traditional morphometry (Ruttner *et al*. 1978; Cornuet & Fresnaye 1989) and the allozyme malate dehydrogenase (Nielsen *et al*. 1994; Smith & Glenn 1995; Arias *et al*. 2006), which have been reported to exhibit a latitudinal gradient along the Mediterranean coast of the Iberian Peninsula.

Contrasting with neutral loci, geographical partitioning recovered by directional outliers argues for selection as an underlying force shaping a latitudinal gradient, as suggested by others (Franck *et al.* 1998; Garnery *et al*. 1998a,b). The spatial analysis identified precipitation, and to a lesser extent temperature as associated with the north–south (latitudinal) selection (Table V - 1). Precipitation was associated along with latitude for nine outlier SNPs, while mean and maximum temperatures were also associated with the precipitation and latitude for four of these nine. What

was not expected was that several directional outliers were found even more strongly correlated with longitude than latitude (see AMOVA and matSAM analysis). Further, while matSAM identified precipitation associated with both longitude and latitude, insolation was identified as uniquely associated with longitude. Insolation is thus identified as a novel parameter shaping population structure east to west. The observation of east–west outlier structuring further adds to the complexity of Iberian honey bees suggesting that the Atlantic portion of the Iberian Peninsula harbours an additional important component of the Iberian honey bee diversity

Concluding remarks

In this study, genome-wide scans using 383 polymorphic SNP loci were conducted in searching for the footprints of selection in the environmentally heterogeneous Iberian honey bee range. A combination of methods was employed leading to the detection of several SNPs marking genomic regions that are promising candidates for adaptation. SNP loci exhibiting strong outlier behaviour were cross-validated by at least three conceptually different F_{ST}-based methods, and a subset was found to be associated with environmental variables that may be causal or correlated selection pressures.

Further support for the role of selection in shaping variation in Iberian honey bees was provided by the functional relevance of the genes carrying outlier SNPs. This is particularly compelling when independent SNPs are mapped to genes that are functionally related. While genome-wide scans provide a powerful way of highlighting candidate genes for selection, experimental support for selection from functional and expression studies (Schluter *et al.* 2010; Renaut *et al*. 2011; Riveron *et al*. 2013) and also indirect evidence from sequence variation analysis (Low *et al*. 2007; Wood *et al*. 2008; Kent *et al*. 2011; Oliveira *et al*. 2012) are ultimately required to make causal inferences about the molecular basis of adaptation of Iberian honey bees. Nevertheless, the approach pursued here enabled, for the first time, a more rigorous consideration of selection as the underlying cause of diversity patterns in Iberian honey bees, whereas in previous studies (Franck *et al.* 1998; Garnery *et al*. 1998a,b), evocation of this evolutionary force to explain clinal variation in the Iberian Peninsula was speculative. Furthermore, our findings represent an important first step towards the identification of polymorphisms implicated in local adaptation and possibly in response to recent human-mediated environmental changes.

References

Ai H, Huang L, Ren J (2013) Genetic diversity, linkage disequilibrium and selection signatures in chinese and Western pigs revealed by genome-wide SNP markers. *PLoS ONE,* **8**, e56001.

Allendorf FW, Luikart G (2007) *Conservation and the Genetics of Populations*, 1st edn. Blackwell Publishing.

Antão T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: a workbench to detect molecular adaptation based on a F_{ST}-outlier method. *BMC Bioinformatics,* **9**, 323.

Arias MC, Rinderer TE, Sheppard WS (2006) Further characterization of honey bees from the Iberian Peninsula by allozyme, morphometric and mtDNA haplotype analyses. *Journal of Apicultural Research,* **45**, 188-196.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology,* **13**, 969-980.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society B-Biological Sciences,* **263**, 1619-1626.

Behura SK, Severson DW (2012) Comparative analysis of codon usage bias and codon context patterns between dipteran and hymenopteran sequenced genomes. *PLoS ONE,* **7**, e43111.

Belyaeva OV, Lee SA, Kolupaev OV, Kedishvili NY (2009) Identification and characterization of retinoid-active short-chain dehydrogenases/reductases in *Drosophila melanogaster*. *Biochimica Et Biophysica Acta-General Subjects,* **1790**, 1266-1273.

Beye M, Gattermeier I, Hasselmann M *et al.* (2006) Exceptionally high levels of recombination across the honey bee genome. *Genome Research,* **16**, 1339-1344.

Blacquière T, Smagghe G, van Gestel CAM, Mommaerts V (2012) Neonicotinoids in bees: a review on concentrations, side-effects and risk assessment. *Ecotoxicology,* **21**, 1581-1581.

Bonin A, Taberlet P, Miaud C, Pompanon F (2006) Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Molecular Biology and Evolution,* **23**, 773-783.

Buchon N, Broderick NA, Poidevin M, Pradervand S, Lemaitre B (2009) *Drosophila* intestinal response to bacterial infection: activation of host defense and stem cell proliferation. *Cell Host and Microbe,* **5**, 200-211.

Buckley J, Butlin RK, Bridle JR (2012) Evidence for evolutionary change associated with the recent range expansion of the British butterfly, *Aricia agestis*, in response to climate change. *Molecular Ecology,* **21**, 267-280.

Bull DL, Whitten CJ (1972) Factors influencing organophosphorus insecticide resistance in tobacco budworms. *Journal of Agricultural and Food Chemistry,* **20**, 561-&.

Bull JC, Ryabov EV, Prince G *et al.* (2012) A strong immune response in young adult honeybees masks their increased susceptibility to infection compared to older bees. *PLoS Pathogens,* **8**, e1003083.

Bunce JA, Isbell LA, Neitz M *et al.*(2011) Characterization of opsin gene alleles affecting color vision in a wild population of titi monkeys (*Callicebus brunneus*). *American Journal of Primatology,* **73**, 189-196.

Burrill RM, Dietz A (1981) The response of honey bees to variations in solar radiation and temperature. *Apidologie,* **12**, 319-328.

Cafferty P, Yu L, Long H, Rao Y (2006) Semaphorin-1a functions as a guidance receptor in the *Drosophila* visual system. *Journal of Neuroscience,* **26**, 3999-4003.

Camps M, Herman A, Loh E, Loeb LA (2007) Genetic constraints on protein evolution. *Critical Reviews in Biochemistry and Molecular Biology,* **42**, 313-326.

Cánovas F, De la Rúa P, Serrano J, Galián J (2008) Geographical patterns of mitochondrial DNA variation in *Apis mellifera iberiensis* (Hymenoptera : Apidae). *Journal of Zoological Systematics and Evolutionary Research,* **46**, 24-30.

Cánovas F, De la Rúa P, Serrano J, Galián J (2011) Microsatellite variability reveals beekeeping influences on Iberian honeybee populations. *Apidologie,* **42**, 235-251.

Claudianos C, Ranson H, Johnson RM *et al.* (2006) A deficit of detoxification enzymes: pesticide sensitivity and environmental response in the honeybee. *Insect Molecular Biology,* **15**, 615-636.

Cooke GM, Chao NL, Beheregaray LB (2012) Divergent natural selection with gene flow along major environmental gradients in Amazonia: insights from genome scans, population genetics and phylogeography of the characin fish *Triportheus albus*. *Molecular Ecology,* **21**, 2410-2427.

Coop G, Witonsky D, Di Rienzo A, Pritchard JK (2010) Using Environmental Correlations to Identify Loci Underlying Local Adaptation. *Genetics,* **185**, 1411-1423.

Cornuet JM, Fresnaye J (1989) Biometrical study of honey bee populations from Spain and Portugal. *Apidologie,* **20**, 93-101.

Cox-Foster DL, Conlan S, Holmes EC *et al.* (2007) A metagenomic survey of microbes in honey bee colony collapse disorder. *Science,* **318**, 283-287.

De Jong MA, Collins S, Beldade P, Brakefield PM, Zwaan BJ (2013) Footprints of selection in wild populations of *Bicyclus anynana* along a latitudinal cline. *Molecular Ecology* **22**, 341-353.

Dietemann V, Pirk CWW, Crewe R (2009) Is there a need for conservation of honeybees in Africa? *Apidologie,* **40**, 285-295.

Dussaubat C, Brunet JL, Higes M *et al.* (2012) Gut pathology and responses to the microsporidium *Nosema ceranae* in the honey bee *Apis mellifera*. *PLoS ONE,* **7**, e37017.

Earl JB, Britt SG (2006) Expression of *Drosophila* rhodopsins during photoreceptor cell differentiation: Insights into R7 and R8 cell subtype commitment. *Gene Expression Patterns,* **6**, 687-694.

Engel MS (1999) The taxonomy of recent and fossil honey bees (Hymenoptera : Apidae; *Apis*). *Journal of Hymenoptera Research,* **8**, 165-196.

Evans JD (2006) Beepath: An ordered quantitative-PCR array for exploring honey bee immunity and disease. *Journal of Invertebrate Pathology,* **93**, 135-139.

Evans JD, Spivak M (2010) Socialized medicine: Individual and communal disease barriers in honey bees. *Journal of Invertebrate Pathology,* **103**, S62-S72.

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity (Edinb),* **103**, 285-298.

Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources,* **10**, 564-567.

Feyereisen R (1999) Insect P450 Enzymes. *Annual Review of Entomology,* **44**, 507–533.

Ffrench-Constant RH, Daborn PJ, Le Goff G (2004) The genetics and genomics of insecticide resistance. *Trends in Genetics,* **20**, 163-170.

Foll M, Gaggiotti O (2006) Identifying the environmental factors that determine the genetic structure of populations. *Genetics,* **174**, 875-891.

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics,* **180**, 977-993.

Franck P, Garnery L, Solignac M, Cornuet JM (1998) The origin of west European subspecies of honeybees (*Apis mellifera*): new insights from microsatellite and mitochondrial data. *Evolution,* **52**, 1119-1134.

Garnery L, Franck P, Baudry E *et al.* (1998a) Genetic diversity of the west European honey bee (*Apis mellifera mellifera* and *A. m. iberica*). I. Mitochondrial DNA. *Genetics Selection Evolution,* **30**, S31-S47.

Garnery L, Franck P, Baudry E *et al.* (1998b) Genetic diversity of the west European honey bee (*Apis mellifera mellifera* and *A. m. iberica*). II. Microsatellite loci. *Genetics Selection Evolution,* **30**, S49-S74.

Garnery L, Mosshine EH, Oldroyd B, P., Cornuet JM (1995) Mitochondrial DNA variation in Moroccan and Spanish honey bee populations. *Molecular Ecology,* **4**, 465-471.

Gisder S, Hedtke K, Mockel N *et al.* (2010) Five-year cohort study of *Nosema spp*. in Germany: Does climate shape virulence and assertiveness of *Nosema ceranae*? *Applied and Environmental Microbiology,* **76**, 3032-3038.

Godenschwege TA, Hu HL, Shan-Crofts X, Goodman CS, Murphey RK (2002) Bi-directional signaling by Semaphorin 1a during central synapse formation in *Drosophila*. *Nature Neuroscience,* **5**, 1294-1301.

Gomez-Uchida D, Seeb JE, Smith MJ *et al.* (2011) Single nucleotide polymorphisms unravel hierarchical divergence and signatures of selection among Alaskan sockeye salmon (*Oncorhynchus nerka*) populations. *BMC Evolutionary Biology,* **11**, 48.

Hardstone MC, Scott JG (2010) Is *Apis mellifera* more sensitive to insecticides than other insects? *Pest Management Science,* **66**, 1171-1180.

Helyar SJ, Hemmer-Hansen J, Bekkevold D *et al.* (2011) Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Molecular Ecology Resources,* **11**, 123-136.

Heymann Y, Steenmans C, G Croissille G, Bossard M (1994) CORINE Land Cover. Office for Official Publications of the European Communities, Luxembourg.

Higes M, Martin-Hernandez R, Garrido-Bailon E *et al.* (2009) Honeybee colony collapse due to *Nosema ceranae* in professional apiaries. *Environmental Microbiology Reports,* **1**, 110-113.

Higes M, Martin-Hernandez R, Meana A (2010) *Nosema ceranae* in Europe: an emergent type C nosemosis. *Apidologie,* **41**, 375-392.

Hollox EJ, Armour JAL (2008) Directional and balancing selection in human beta-defensins. *BMC Evolutionary Biology,* **8**, 113.

Jaffe R, Dietemann V, Allsopp MH *et al.* (2010) Estimating the density of honeybee colonies across their natural range to fill the gap in pollinator decline censuses. *Conservation Biology,* **24**, 583-593.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics,* **23**, 1801-1806.

Jiggins FM, Hurst GDD (2003) The evolution of parasite recognition genes in the innate immune system: Purifying selection on *Drosophila melanogaster* peptidoglycan recognition proteins. *Journal of Molecular Evolution,* **57**, 598-605.

Jin LH, Shim J, Yoon JS *et al.* (2008) Identification and functional analysis of antifungal immune response genes in *Drosophila*. *PLoS Pathogens,* **4**, e1000168.

Johnson RM, Wen Z, Schuler MA, Berenbaum MR (2006) Mediation of pyrethroid insecticide toxicity to honey bees (Hymenoptera : Apidae) by cytochrome P450 monooxygenases. *Journal of Economic Entomology,* **99**, 1046-1050.

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics,* **24**, 1403-1405.

Joost S, Bonin A, Bruford MW *et al.* (2007) A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Molecular Ecology,* **16**, 3955-3969.

Joost S, Kalbermatten M, Bonin A (2008) Spatial analysis method(SAM): a software tool combining molecular and environmental data to identify candidate loci for selection. *Molecular Ecology Resources,* **8**, 957-960.

Kent CF, Issa A, Bunting AC, Zayed A (2011) Adaptive evolution of a key gene affecting queen and worker traits in the honey bee, *Apis mellifera*. *Molecular Ecology* **20**, 5226-5235.

Kim TS, Maeda A, Maeda T *et al.*(2005) Delayed dark adaptation in 11-cis-retinol dehydrogenase-deficient mice - A role of RDH11 in visual processes in vivo. *Journal of Biological Chemistry,* **280**, 8694-8704.

Kinel-Tahan Y, Weiss H, Dgany O, Levine A, Wides R (2007) *Drosophila* odz gene is required for multiple cell types in the compound retina. *Developmental Dynamics,* **236**, 2541-2554.

Kurucz E, Markus R, Zsamboki J *et al.* (2007) Nimrod, a putative phagocytosis receptor with EGF repeats in *Drosophila* plasmatocytes. *Current Biology,* **17**, 649-654.

Ladd AN, Cooper TA (2002) Finding signals that regulate alternative splicing in the postgenomic era. *Genome Biology,* **3**, reviews0008.

Lehtonen PK, Laaksonen T, Artemyev AV *et al.* (2012) Candidate genes for colour and vision exhibit signals of selection across the pied flycatcher (*Ficedula hypoleuca*) breeding range. *Heredity,* **108**, 431-440.

Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics,* **74**, 175-195.

Li X, Schuler MA, Berenbaum MR (2007) Molecular mechanisms of metabolic resistance to synthetic and natural xenobiotics. *Annual Review of Entomology*, **52**, 231-253.

Limborg MT, Helyar SJ, de Bruyn M *et al.* (2012) Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (*Clupea harengus*). *Molecular Ecology,* **21**, 3686-3703.

Low WY, Ng HL, Morton CJ *et al.* (2007) Molecular evolution of glutathione S-transferases in the genus *Drosophila*. *Genetics,* **177**, 1363-1375.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews. Genetics,* **4**, 981-994.

Manel S, Poncet BN, Legendre P, Gugerli F, Holderegger R (2010) Common factors drive adaptive genetic variation at different spatial scales in *Arabis alpina*. *Molecular Ecology,* **19**, 3824-3835.

Mao W, Rupasinghe SG, Johnson RM *et al.* (2009) Quercetin-metabolizing CYP6AS enzymes of the pollinator *Apis mellifera* (Hymenoptera: Apidae). *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology,* **154**, 427-434.

Mao W, Schuler MA, Berenbaum MR (2011) CYP9Q-mediated detoxification of acaricides in the honey bee (*Apis mellifera*). *Proceedings of the National Academy of Sciences of the United States of America,* **108**, 12657-12662.

Meixner MD, Leta MA, Koeniger N, Fuchs S (2011) The honey bees of Ethiopia represent a new subspecies of *Apis mellifera-Apis mellifera simensis* n. ssp. *Apidologie,* **42**, 425-437.

Menzel R, Müller U (1996) Learning and memory in honey bees: from behavior to neural substrates. *Annual Review of Neuroscience,* **19**, 379–404.

Miguel I, Baylac M, Iriondo M *et al.* (2011) Both geometric morphometric and microsatellite data consistently support the differentiation of the *Apis mellifera* M evolutionary branch. *Apidologie,* **42**, 150-161.

Miguel I, Iriondo M, Garnery L, Sheppard WS, Estonba A (2007) Gene flow within the M evolutionary lineage of *Apis mellifera*: role of the Pyrenees, isolation by distance and post-glacial re-colonization routes in the western Europe. *Apidologie,* **38**, 141-155.

Moritz RFA, Kraus FB, Kryger P, Crewe RM (2007) The size of wild honeybee populations (*Apis mellifera*) and its implications for the conservation of honeybees. *Journal of Insect Conservation,* **11**, 391-397.

Narum SR, Hess JE (2011) Comparison of F_{ST} outlier tests for SNP loci under selection. *Molecular Ecology Resources,* **11**, 184-194.

Navajas M, Migeon A, Alaux C *et al.*(2008) Differential gene expression of the honey bee *Apis mellifera* associated with *Varroa destructor* infection. *BMC Genomics,* **9**, 301.

Nazzi F, Brown SP, Annoscia D *et al.* (2012) Synergistic parasite-pathogen interactions mediated by host immunity can drive the collapse of honeybee colonies. *PLoS Pathogens,* **8**, e1002735.

Nielsen D, Page RE, Jr., Crosland MW (1994) Clinal variation and selection of MDH allozymes in honey bee populations. *Experientia,* **50**, 867-871.

Nielsen EE, Hemmer-Hansen J, Poulsen NA *et al.*(2009) Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). *BMC Evolutionary Biology,* **9**, 276.

Nielsen R (2005) Molecular signatures of natural selection. *Annual Review of Genetics,* **39**, 197-218.

Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology,* **18**, 375-402.

Nunes VL, Beaumont MA, Butlin RK, Paulo OS (2011) Multiple approaches to detect outliers in a genome scan for selection in ocellated lizards (*Lacerta lepida*) along an environmental gradient. *Molecular Ecology,* **20**, 193-205.

Oliveira TYK, Harris EE, Meyer D, Jue CK, Silva WA (2012) Molecular evolution of a malaria resistance gene (DARC) in primates. *Immunogenetics,* **64**, 497-505.

Osborne JL (2012) ECOLOGY Bumblebees and pesticides. *Nature,* **491**, 43-45.

Pérez-Figueroa A, Garcia-Pereira MJ, Saura M, Rolan-Alvarez E, Caballero A (2010) Comparing three different methods to detect selective loci using dominant markers. *Journal of Evolutionary Biology,* **23**, 2267-2276.

Pilling ED, Bromleychallenor KAC, Walker CH, Jepson PC (1995) Mechanism of synergism between the pyrethroid insecticide lambda-cyhalothrin and the imidazole fungicide prochloraz, in the honeybee (*Apis mellifera* L.). *Pesticide Biochemistry and Physiology,* **51**, 1-11.

Pinto MA, Henriques D, Neto M *et al.* (2013) Maternal diversity patterns of Ibero-Atlantic populations reveal further complexity of Iberian honey bees. *Apidologie,* **44**, 430–439.

Potts SG, Biesmeijer JC, Kremen C *et al.* (2010) Global pollinator declines: trends, impacts and drivers. *Trends in Ecology & Evolution,* **25**, 345-353.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics,* **155**, 945–959.

Prunier J, Laroche J, Beaulieu J, Bousquet J (2011) Scanning the genome for gene SNPs related to climate adaptation and estimating selection at the molecular level in boreal black spruce. *Molecular Ecology,* **20**, 1702-1716.

Renaut S, Nolte AW, Rogers SM, Derome N, Bernatchez L (2011) SNP signatures of selection on standing genetic variation and their association with adaptive phenotypes along gradients of ecological speciation in lake whitefish species pairs (*Coregonus spp.*). *Molecular Ecology,* **20**, 545-559.

Riveron JM, Irving H, Ndula M *et al.* (2013) Directionally selected cytochrome P450 alleles are driving the spread of pyrethroid resistance in the major malaria vector *Anopheles funestus*. *Proceedings of the National Academy of Sciences of the United States of America,* **110**, 252-257.

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes,* **4**, 137-138.

Ruttner F (1988) *Biogeography and taxonomy of honey bees* Springer, Berlin.

Ruttner F, Tassencourt L, Louveaux J (1978) Biometrical-statistical analysis of the geographic variability of *Apis mellifera* L. *Apidologie,* **9**, 363-381.

Sackton TB, Lazzaro BP, Schlenke TA *et al.*(2007) Dynamic evolution of the innate immune system in *Drosophila*. *Nature Genetics,* **39**, 1461-1468.

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning*, 2nd edn. Cold Spring Harbor Laboratory Press, New York.

Schlötterer C (2002) A microsatellite-based multilocus screen for the identification of local selective sweeps. *Genetics,* **160**, 753-763.

Schluter D, Marchinko KB, Barrett RDH, Rogers SM (2010) Natural selection and the genetics of adaptation in threespine stickleback. *Philosophical Transactions of the Royal Society B-Biological Sciences,* **365**, 2479-2486.

Scott JG, Wen ZM (2001) Cytochromes P450 of insects: the tip of the iceberg. *Pest Management Science,* **57**, 958-967.

Sheppard WS, Meixner MD (2003) *Apis mellifera pomonella*, a new honey bee subspecies from Central Asia. *Apidologie,* **34**, 367-375.

Shikano T, Ramadevi J, Merilä J (2010) Identification of local- and habitat-dependent selection: scanning functionally important genes in nine-spined sticklebacks (*Pungitius pungitius*). *Molecular Biology and Evolution,* **27**, 2775-2789.

Shimada Y, Shikano T, Merilä J (2011) A high incidence of selection on physiologically important genes in the three-spined stickleback, *Gasterosteus aculeatus*. *Molecular Biology and Evolution,* **28**, 181-193.

Smith DR, Glenn TC (1995) Allozyme polymorphisms in spanish honeybees (*Apis mellifera iberica*). *Journal of Heredity,* **86**, 12-16.

Smith DR, Palopoli MF, Taylor BR *et al.* (1991) Geographical overlap of two mitochondrial genomes in Spanish honeybees (*Apis mellifera iberica*). *Journal of Heredity,* **82**, 96-100.

Somogyi K, Sipos B, Penzes Z *et al.* (2008) Evolution of genes and repeats in the Nimrod superfamily. *Molecular Biology and Evolution,* **25**, 2337-2347.

Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology,* **14**, 671-688.

Thioulouse J, Chessel D, Doledec S, Olivier JM (1997) ADE-4: A multivariate analysis and graphical display software. *Statistics and Computing,* **7**, 75-83.

Townson SM, Chang BSW, Salcedo E *et al.* (1998) Honeybee blue-and ultraviolet-sensitive opsins: Cloning, heterologous expression in *Drosophila*, and physiological characterization. *Journal of Neuroscience,* **18**, 2412-2422.

Tsai CJ, Sauna ZE, Kimchi-Sarfaty C *et al.* (2008) Synonymous mutations and ribosome stalling can lead to altered folding pathways and distinct minima. *Journal of Molecular Biology,* **383**, 281-291.

Tsumura Y, Uchiyama K, Moriguchi Y, Ueno S, Ihara-Ujino T (2012) Genome scanning for detecting adaptive genes along environmental gradients in the Japanese conifer, *Cryptomeria japonica*. *Heredity,* **109**, 349-360.

vanEngelsdorp D, Evans JD, Saegerman C *et al.* (2009) Colony collapse disorder: a descriptive study. *PLoS ONE,* **4**, e6481.

vanEngelsdorp D, Meixner MD (2010) A historical review of managed honey bee populations in Europe and the United States and the factors that may affect them. *Journal of Invertebrate Pathology,* **103**, S80-S95.

Vasemägi A, Primmer CR (2005) Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Molecular Ecology,* **14**, 3623-3642.

Vitalis R, Dawson K, Boursot P (2001) Interpretation of variation across marker loci as evidence of selection. *Genetics,* **158**, 1811-1823.

Wang Q, Hasan G, Pikielny CW (1999) Preferential expression of biotransformation enzymes in the olfactory organs of *Drosophila melanogaster*, the antennae. *Journal of Biological Chemistry,* **274**, 10309-10315.

Wang T, Chen GP, Zan QJ, Wang CB, Su YJ (2012) AFLP genome scan to detect genetic structure and candidate loci under selection for local adaptation of the invasive weed *Mikania micrantha*. *PLoS ONE,* **7**, e41310.

Wang XY, Wang T, Jiao YC, von Lintig J, Montell C (2010) Requirement for an enzymatic visual cycle in *Drosophila*. *Current Biology,* **20**, 93-102.

Watson FL, Puttmann-Holgado R, Thomas F *et al.* (2005) Extensive diversity of Ig-superfamily proteins in the immune system of insects. *Science,* **309**, 1874-1878.

Whitfield CW, Behura SK *et al.* (2006) Thrice out of Africa: ancient and recent expansions of the honey bee, *Apis mellifera. Science,* **314**, 642-645.

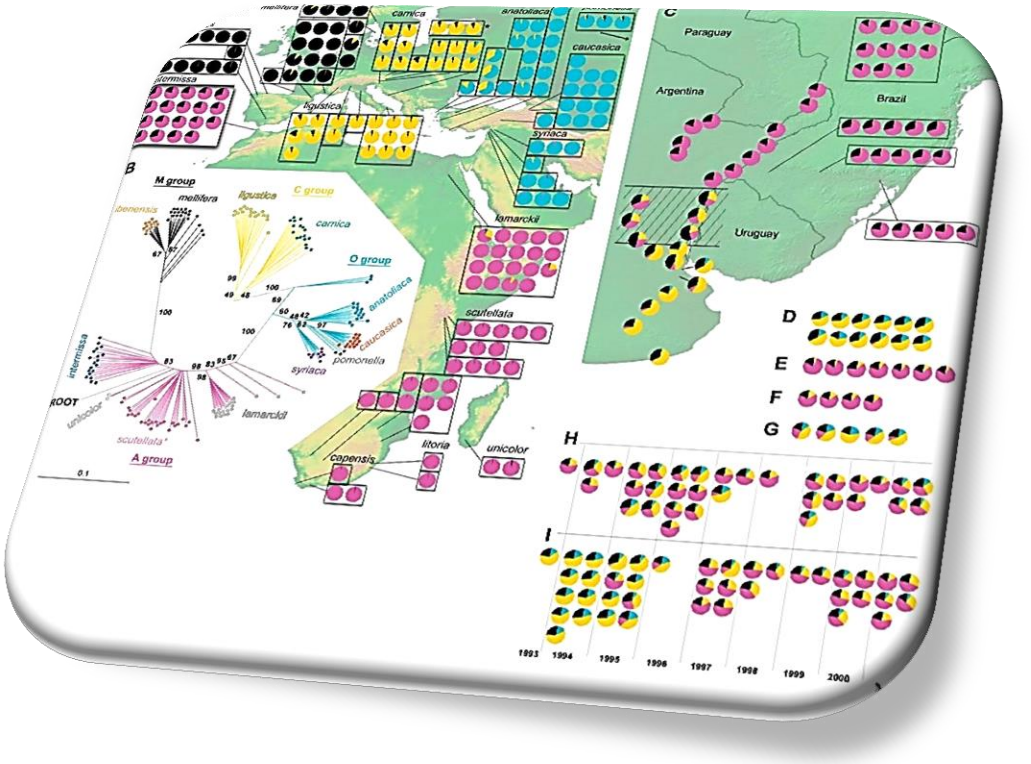
Wilding CS, Butlin RK, Grahame J (2001) Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. *Journal of Evolutionary Biology,* **14**, 611-619.

Winston M (1987) *The biology of the honey bee*, London.

Wood HM, Grahame JW, Humphray S, Rogers J, Butlin RK (2008) Sequence differentiation in regions identified by a genome scan for local adaptation. *Molecular Ecology,* **17**, 3123-3135.

Yu SJ (2002) Biochemical characteristics of microsomal and cytosolic glutathione S-transferases in larvae of the fall armyworm, *Spodoptera frugiperda* (J. E. Smith). *Pesticide Biochemistry and Physiology,* **72**, 100-110.

Zayed A, Whitfield CW (2008) A genome-wide signature of positive selection in ancient and recent invasive expansions of the honey bee *Apis mellifera*. *Proceedings of the National Academy of Sciences of the United States of America,* **105**, 3421-3426.



Chapter VI

Revisiting the Iberian honey bee (*Apis mellifera iberiensis*) contact zone: maternal and genome-wide nuclear variations provide support for secondary contact from historical refugia

The work presented in this chapter has been published:

Chávez-Galarza J, Henriques D, Johnston JS, Carneiro M, Rufino J, Patton JC, Pinto MA (2015) Revisiting the Iberian honey bee (*Apis mellifera iberiensis*) contact zone: maternal and genome-wide nuclear variations provide support for secondary contact from historical refugia. **Molecular Ecology**, 24: 2973-2992

Abstract

Dissecting diversity patterns of organisms endemic to Iberia has been truly challenging for a variety of taxa, and the Iberian honey bee is no exception. Surveys of genetic variation in the Iberian honey bee are among the most extensive for any honey bee subspecies. From these, differential and complex patterns of diversity have emerged, which have yet to be fully resolved. Here, we used a genome-wide data set of 309 neutrally tested single nucleotide polymorphisms (SNPs), scattered across the 16 honey bee chromosomes, which were genotyped in 711 haploid males. These SNPs were analysed along with an intergenic locus of the mtDNA, to reveal historical patterns of population structure across the entire range of the Iberian honey bee. Overall, patterns of population structure inferred from nuclear loci by multiple clustering approaches and geographic cline analysis were consistent with two major clusters forming a well-defined cline that bisects Iberia along a northeastern–southwestern axis, a pattern that remarkably parallels that of the mtDNA. While a mechanism of primary intergradation or isolation by distance could explain the observed clinal variation, our results are more consistent with an alternative model of secondary contact between divergent populations previously isolated in glacial refugia, as proposed for a growing list of other Iberian taxa. Despite current intense honey bee management, human-mediated processes have seemingly played a minor role in shaping Iberian honey bee genetic structure. This study highlights the complexity of the Iberian honey bee patterns and reinforces the importance of Iberia as a reservoir of *Apis mellifera* diversity.

Keywords: *Apis mellifera iberiensis*, geographic cline analysis, honey bee, Iberia, secondary contact, SNPs, sPCA, structure

Introduction

Clinal patterns in gene frequencies can be generated by random genetic drift under an isolation-by-distance scenario. Alternatively, clinal variation may be shaped by selection acting within a continuous population (primary intergradation) or, more frequently, may originate from contact between populations that diverged in allopatry (secondary contact). Distinguishing primary intergradation from secondary contact can, however, be a difficult undertaking because both processes may generate similar patterns of genetic variation (Endler 1977; Barton & Hewitt 1985, 1989). Population genomics provides a suitable framework in which to more effectively unravel such levels of complexity. In population genomics, outlier tests are applied to genome-wide sampling of multiple populations to dissect out adaptive variation, leaving a background of neutral and near-neutral variation (Luikart *et al*. 2003). Cline analysis can then help reveal whether dissected patterns of variation originated from secondary contact or primary intergradation. If multiple coincident clines are identified (Endler 1977; Barton & Hewitt 1985) and these clines reflect changes in neutral loci, there is strong support for recent secondary contact (Durrett *et al*. 2000). Unless many independent loci respond similarly to a single environmental gradient or mosaic, clinal patterns of neutral variation and multiple coincident clines are not expected when primary intergradation is the leading process shaping variation (Durrett *et al.* 2000).

The Iberian Peninsula provides one of the most interesting settings in Europe for studying contact zones. High geological, physiographical and climatic complexity and diversity, together with isolation from Europe and proximity to Africa (especially at the Strait of Gibraltar), made this southernmost European region an important refuge during the Quaternary glaciations (reviewed by Hewitt 2000) and a bridge, for the more vagile organisms, between the two continents (Carranza *et al*. 2004; Cosson *et al*. 2005; Guillaumet *et al*. 2006; Whitfield *et al*. 2006; Wallberg *et al*. 2014). These features made Iberia not only a place of divergence during periods of isolation but also a contact zone during periods of expansion as reported for a wide array of plant and animal taxa (extensively reviewed by Weiss & Ferrand 2007), including the focal organism of this study: the Iberian honey bee, *Apis mellifera iberiensis*.

Disentangling diversity patterns in populations that have possibly experienced recurrent cycles of contraction, expansion, admixture, and adaptation, typical of long-term glacial refugia, is a challenging endeavour. Contemporary human-mediated processes, which in the case of the honey bee may involve movement of colonies within (transhumance) and between lineages (introduction of commercial queens), selective breeding, and accidental introductions of exotic pests and

diseases, may further complicate this effort by erasing or obscuring the genetic signatures imprinted by evolutionary and demographic processes. Fortunately, however, Iberia is the best-studied refugial area in Europe, and common patterns are emerging from comparative phylogeography (Gómez & Lunt 2007) that are of great assistance in elucidating patterns exhibited by the Iberian honey bee. Additionally, the honey bee genome has been sequenced and a SNP panel is available for conducting genome-wide sampling of multiple populations (Whitfield *et al*. 2006; Chávez-Galarza *et al*. 2013).

The honey bee native range spans Africa, Europe and the Middle East where it evolved into 30 subspecies (Ruttner 1988; Engel 1999; Sheppard & Meixner 2003; Meixner *et al*. 2011). This vast amount of variation has been grouped into four (western European, M; eastern European, C; African, A; Middle Eastern, O) largely parapatric evolutionary lineages (Ruttner 1988; Garnery *et al*. 1992; Whitfield *et al*. 2006; Wallberg *et al*. 2014), with contact zones identified in Italy (Franck *et al*. 2000), Turkey (Kandemir *et al*. 2006), Libya (Shaibi *et al*. 2009), and Iberia (Smith *et al*. 1991; Franck *et al*. 1998; Garnery *et al*. 1998a; Arias *et al*. 2006; Miguel *et al*. 2007; Cánovas *et al*. 2008). Among these, the Iberian contact zone formed by A and M lineages has received the greatest attention with numerous studies using a wide array of markers ranging from morphology (Cornuet & Fresnaye 1989; Arias *et al.* 2006; Miguel *et al.* 2011), allozymes (Smith & Glenn 1995; Arias *et al*. 2006), mitochondrial DNA (Smith *et al*. 1991; Garnery *et al*. 1995, 1998a; Franck *et al*. 1998; De la Rúa *et al*. 2001, 2004, 2005; Arias *et al*. 2006; Miguel *et al*. 2007; Cánovas *et al*. 2008; Pinto *et al*. 2012, 2013), microsatellites (Franck *et al*. 1998; Garnery *et al*. 1998b; De la Rúa *et al*. 2002, 2003; Miguel *et al*. 2007, 2011; Cánovas *et al*. 2011) to SNPs (Chávez-Galarza *et al.* 2013).

Differential and complex diversity patterns emerged from the numerous biparental and maternal surveys of Iberian honey bees and the underlying processes shaping genetic variation remain controversial. Arguments based on selection, demography, and contemporary human-mediated processes have been favoured by different authors. Morphology (Ruttner *et al*. 1978; Cornuet & Fresnaye 1989) and the allozyme malate dehydrogenase (Nielsen *et al*. 1994; Smith & Glenn 1995; Arias *et al*. 2006) exhibited a smooth latitudinal cline extending from North Africa to France supporting a hypothesis of primary intergradation (Ruttner *et al*. 1978). In contrast, the abrupt transition from highly divergent M mitotypes in the northeastern half of Iberia to A mitotypes in the southwestern half (Garnery *et al*. 1995; Franck *et al*. 1998; Arias *et al*. 2006; Miguel *et al*. 2007; Cánovas *et al*. 2008) was more compatible with secondary contact (Smith *et al*. 1991). To

complicate matters further, microsatellites did not capture the signal of a contact zone in Iberia (Franck *et al*. 1998; Garnery *et al*. 1998b; Cánovas *et al*. 2011; Miguel *et al*. 2011), but detected instead a sharp disruption between Iberian and northern African populations (Franck *et al*. 1998). This latter finding prompted a third hypothesis that explained occurrence of A mitotypes in Iberia by human-assisted introductions of African colonies during Muslim occupation, with selection acting to maintain the morphological, allozymic and maternal clines (Franck *et al*. 1998). The hypothesis of selection as the driving force shaping the Iberian cline was recently addressed in a genome-wide SNP scan conducted in a fine-scale sample that covered the entire Iberian honey bee range (Chávez-Galarza *et al*. 2013). This study detected signatures of selection in the Iberian honey bee genome, suggesting that this evolutionary force has had an important role in structuring Iberian honey bee diversity.

Here, we built from those findings to provide, at both geographic and genomic levels, the most comprehensive characterization of the Iberian honey bee diversity patterns performed until now. We employed multiple clustering approaches and cline analysis to examine the genome-wide SNP data set using a population-genomics framework. After analysing the patterns of variation generated by the complete SNP data set, we removed any SNPs putatively associated with selection identified by Chávez-Galarza *et al.* (2013) and then used concurrently a mtDNA locus and 309 remaining neutrally tested SNPs to address the following questions: (i) How effective are SNPs in capturing clinal variation? – a signal that microsatellites have failed to detect, (ii) How concordant are the patterns generated by the complete and the neutral SNP data sets? (iii) Do neutral SNPs capture the clinal signal? (iv) Does the mtDNA marker confirm the presence of a cline formed by two highly divergent lineages, as documented by earlier studies? and How concordant are the patterns of neutral and maternal variation? If variation originated from secondary contact, then we expect neutral SNPs to detect clinal patterns and multiple coincident clines. In contrast, coincident neutral clines are not expected if variation originated via primary intergradation. Further evidence for secondary contact will come from comparisons of mtDNA and nuclear DNA. If a maternal cline formed by two highly divergent lineages is observed and this cline is paralleled by nuclear DNA variation, then there is strong support for secondary contact. However, given that Iberian honey bees are managed organisms, it is possible that human-mediated processes have obscured historical patterns. Therefore, we also asked the question: (v) To what extent do contemporary human-mediated forces influence the Iberian honey bee structure?

Methods

Sampling

Sampling in Iberia was conducted in 2010 across three north–south transects (Fig. VI - 1). One transect extended along the Atlantic coast [Atlantic transect (AT)], one through the centre [central transect (CT)], and another along the Mediterranean coast [Mediterranean transect (MT)]. A total of 711 honey bee haploid males were collected in the three transects from 23 sites (AT=8, CT=9; MT=6) representing 237 apiaries and 711 colonies. Samples were stored in absolute ethanol at -20 °C until molecular analysis. Global positioning system (GPS) coordinates were recorded in the field for each apiary. Further details on sampling procedure can be found in Chávez-Galarza *et al.* (2013).

Figure VI - 1. Map of the Iberian Peninsula showing the centroids of the sampling sites at each transect (AT-Atlantic, CT-central, MT-Mediterranean), sample size per site (n), site codes (AT1 to MT6), and the west-east transect (dashed line) traced from Lisbon to Girona for the geographic cline analysis (see Fig. VI - 6). The approximate location of putative inferred refugia for the Iberian honey bee and other Iberian fauna that supports them (adapted from Gomez & Lunt 2007) is also shown.

DNA extraction and marker genotyping

Using a phenol–chloroform–isoamyl alcohol (25:24:1) protocol (Sambrook *et al*. 1989), total DNA was extracted from the thorax of the 711 individuals, each representing a single colony. Analysis of mtDNA was based on the very popular $tRNA^{leu}$ - cox2 intergenic locus. This region was amplified using the primers and PCR conditions detailed elsewhere (Garnery *et al*. 1993). After PCR amplification, the products were sequenced in both directions. Analysis of nuclear DNA was based on SNPs, which were genotyped using Illumina's Bead-Array Technology and the Illumina GoldenGate® Assay with a custom Oligo Pool Assay (Illumina, San Diego, CA, USA) following manufacturer's protocols. Genotype calling was performed using ILLUMINA'S GENOMESTUDIO® data analysis software.

Of the 1536 SNPs included in the GoldenGate array, 383 passed the quality filtering and were polymorphic in *A. m. iberiensis*, using a cut-off criterion of > 0.98 for the most common allele. The 383 SNPs (referred hereafter as the complete SNP data set) were scanned for signatures of selection using four multiple-population F_{ST}-based methods and the spatial method matSAM (Chávez-Galarza *et al*. 2013). The 74 outlier loci detected by at least one of the five methods were removed from the complete data set, leaving 309 neutrally tested SNP loci (referred hereafter as the neutral SNP data set). Chromosomal locations in honey bee linkage groups and a summary of physical distances of the SNPs are shown in Table Sup VI - 1 (Supporting information). Details of SNP genotyping and detection of outlier loci can be found in Chávez-Galarza *et al*. (2013).

Mitochondrial DNA analysis

The sequences produced for the $tRNA^{leu}$ - cox2 intergenic locus were aligned using MEGA version 5.03 (Tamura *et al*. 2011). For the identification of mitotypes, the sequences were compared with those available in GenBank. Each mitotype was then assigned to western European (M), eastern European (C), and African (A) lineages or to an African sublineage (A_I, A_{II}, A_{III}), according to the complex architecture of the $tRNA^{leu}$ - cox2 intergenic region described elsewhere (Garnery *et al*. 1993; Franck et al. 1998, 2001; Rortais *et al*. 2011; Pinto *et al*. 2012). In short, the intergenic region is composed of two elements: the P (size varies between ~53 and 68 bp) and the Q (size varies between ~194 and 196 bp). A combination of point mutations and indels in the P element distinguishes honey bee subspecies from different lineages. The number of Q elements can vary between one and four, although the number of repeats is not lineage specific. Given the highly

variable size of the sequences resulting predominantly from the variable number of Q elements, the sequences were trimmed at the end of the first Q with additional Qs coded as present/absent. Therefore, the ~627-bp sequence fragment analysed herein encompassed the 50 end of the $tRNA^{leu}$ gene, the P element, the first Q element, the coding relative to the other Q elements, and the 30 end of the cox2 gene. Relationships among the sequences were inferred using the median-joining network algorithm (Bandelt *et al*. 1999), as implemented in the program NETWORK version 4.6.1.1 (Fluxus Engineering, Clare, UK; http://www.fluxusengineering.com). The trimmed ~627-bp sequences examined in this study did not diverge from those downloaded from GenBank. Therefore, GenBank sequences were not included in the network analysis.

Estimation of structure by non-spatial approaches

Structure was inferred from the complete and the neutral SNP data sets using two approaches: the Bayesian model-based STRUCTURE and the model-free discriminant analysis of principal components (DAPC). These approaches were used to estimate the proportion of an individual's genome (*Q*) that originated from a given genetic group or cluster.

The Bayesian clustering approach was implemented in STRUCTURE 3.4 (Pritchard *et al*. 2000) for haploid data using the admixture ancestry and correlated allele frequency models run with the unsupervised option. The program was set up for 750 000 Markov chain Monte Carlo (MCMC) iterations after an initial burn-in of 250 000, which was sufficient to reach convergence. Over 20 independent runs for each number of clusters (K), from 1 to 7, were performed to confirm consistency across runs. The Greedy algorithm, implemented in the software CLUMPP 1.1.2 (Jakobsson & Rosenberg 2007), was used to compute the pairwise 'symmetric similarity coefficient' between pairs of runs and to align the 20 runs for each K. The means of the permuted results were plotted using the software DISTRUCT 1.1 (Rosenberg 2004). The optimal K value was determined using Evanno's ΔK (Evanno *et al*. 2005) and Campana's ΔF_{st} methods (Campana *et al*. 2011) in STRUCTURE HARVESTER web v0.6.93 (Earl & Von Holdt 2012) and CORRSIEVE 1.6-8 package (Campana *et al*. 2011), respectively.

The DAPC clustering approach was implemented in ADEGENET 1.3-9 package for R (Jombart 2008). Simulation studies have shown that DAPC performs as well or better than STRUCTURE, particularly under more complex structuring scenarios (Jombart *et al*. 2010; Klaassen *et al*. 2012). DAPC provides a description of the genetic clustering using coefficients of the alleles (loadings) in linear combinations and seeks to maximize between-groups variance and

minimize within-group variance in these loadings (Jombart *et al*. 2010). Successive K-means clustering runs (from 1 to 40) were also incorporated in the analysis to estimate the optimal number of distinct clusters (*K*) based on the Bayesian information criterion (BIC). The optimal *K* value is associated with the lowest BIC value (Jombart *et al*. 2010).

Estimation of structure by spatial approaches

Spatial structure was inferred from the complete and the neutral SNP data sets using two approaches that explicitly incorporate information on geographic coordinates for genotyped individuals: the model-free multivariate spatial principal component analysis (sPCA) and the Bayesian model-based TESS. The sPCA is a modification of PCA which takes into account not only the genetic variance of individuals or populations but also their spatial autocorrelation (measured by Moran's I). This approach disentangles global structures (clines, patches or intermediates) from local structures (strong genetic differences among neighbours), and from random noise (random distribution of allelic frequencies among individuals or populations on a connection network). While global structures display positive spatial autocorrelation (high positive eigenvalue), local structures display negative spatial autocorrelation (high negative eigenvalue) (Jombart *et al*. 2008). The sPCA was performed in ADEGENET using the *K*-nearest neighbours to model the spatial connectivity among individuals. To test for global and local structures, a Monte Carlo test was implemented using 10 000 permutations.

The Bayesian model-based clustering approach implemented by the software TESS (Chen *et al*. 2007) incorporates spatial population models assuming geographic continuity of allele frequencies by including the interaction parameter Ψ, which defines the intensity of two neighbouring individuals belonging to the same genetic cluster. The incorporation of trend surface and a Gaussian autoregressive residual term allows for capturing global and local patterns. The software TESS 2.3.1 was run for haploid data using the convolution admixture model (BYM), correlated allele frequency and a trend degree surface of 1. A Euclidean distance matrix was used to weight the spatial connectivity among individuals. Five runs were carried out at each *K*, from 2 to 7, with 5 000 000 MCMC total sweeps including a burn-in of 1 250 000 sweeps. For each run, the deviance information criterion (DIC) was calculated and the values of all runs were averaged and plotted against *K*. The first lower DIC value represents the optimal *K* for the data. As in STRUCTURE, the software programs CLUMPP and DISTRUCT were used to obtain the average matrix of membership proportions (*Q*) for each *K* and for graphical representation.

Geographic cline analysis

Geographic clines were estimated for mtDNA frequency, individual SNPs frequency, and mean *Q* per sampling site inferred from both the complete and the neutral SNP data sets by STRUCTURE at K = 2. Cline analysis of individual SNPs was performed on a subset of loci with an absolute allele frequency difference > 0.2. The rationale behind this choice was that the SNP panel used in this study was ascertained from the reference genome of *A. mellifera* (sequenced from the North American DH4 strain, which was primarily *A. m. ligustica*, a subspecies belonging to the C lineage) and genome sequence traces of Africanized honey bees (largely the African *A. m. scutellata* admixed with the genomes of *A. m. ligustica* and the M-lineage *A. m. mellifera*), and is thus underrepresented for diagnostic SNPs and SNPs with large allele frequency differences between the two maternal lineages identified in Iberian honey bees. Using a subset of loci with larger allele frequency difference between the groups, we expected to increase the ancestry information of individual loci for cline analysis. Of the 383 SNPs, 33 (17 neutral and 16 under selection, as identified by Chávez-Galarza *et al*. 2013) conformed to the frequency criterion.

Sampling sites were arranged along a transect beginning at the westernmost location (Lisbon, Portugal) and ending at the easternmost location (Girona, Spain) (Fig. VI - 1). Each sampling site was assigned a distance along this transect, which corresponded to the shortest straight-line distance between it and Lisbon, calculated using the 'harversine' approach (www.movable-type.co.uk/scripts/latlong.html). The cline shape was modelled using the package HZAR v2.5 (Derryberry *et al*. 2014), which fits allele frequency data to equilibrium geographic cline models (Szymura & Barton 1986, 1991; Barton & Gale 1993; Gay *et al*. 2008) using the Metropolis–Hastings Markov chain Monte Carlo algorithm. The following cline shape parameters were estimated: centre (c, distance from sampling location), width (w, 1/maximum slope), delta (δ, distance between the centre and the tail) and tau (τ, slope of the tail). The allele frequencies at the top and bottom of the cline (*Pmin* and *Pmax*) were either fixed or free to vary. Three sets of five cline models were fitted: model set 1 had no scaling (*Pmin* = 0, *Pmax* = 1), model set 2 had fixed scaling (*Pmin* = observed minimum, *Pmax* = observed maximum), and model set 3 allowed *Pmin* and *Pmax* to vary. Within each model set, scaling and tails were fixed or free to vary. These models were compared to a null model of no clinal transition using the Akaike Information Criterion corrected (AICc). The best-fitting model had the lowest AICc value. To evaluate coincidence among cline centre positions and concordance among cline widths, the composite likelihood method (Phillips *et al*. 2004) was used. Likelihood profiles were constructed for both c and w to compare

alternative hypotheses across loci: H1, all loci are characterized by statistically indistinguishable c and w values and are likely to share a common c/w; and H2, each locus has its own independent c and w values. Composite log-likelihood profiles were constructed by summing log-likelihood (ML) profiles for all individual SNP loci ML(H1). This composite log-likelihood profile was compared to the sum of all maximum-likelihood estimates for individual SNP loci ML(H2) using a likelihood ratio test (LRT). If the clines of individual SNP loci coincide and have the same c/w values, ML(H1) is not significantly different from ML (H2) (ML = ML(H2) – ML(H1) ≈ 0). Conversely, if the clines do not coincide, ML(H1) is significantly smaller than ML(H2) (ML > 0). The significance of any difference of ML(H1) and ML(H2) was determined using a chi-square test with n−1 degrees of freedom (α = 0.05). This approach was similarly employed to evaluate coincidence and concordance of mtDNA and both SNP data sets.

Linkage disequilibrium and genetic diversity

Linkage disequilibrium (LD) between all pairs of neutral SNPs was estimated using the statistic r^2 (Hill & Robertson 1968), as implemented by the software DNASP 5.10.1 (Librado & Rozas 2009). Significant LD was identified at the 5% level using Fisher's exact test. Unbiased haploid genetic diversity (*uh*) for the neutral SNPs was calculated using the program GENALEX 6.5 (Peakall & Smouse 2012). Values of LD and *uh* were calculated for each sampling site and then projected along the Lisbon–Girona transect (Fig. VI - 1).

Statistical tests

Differences in individuals' *Q*s between clustering approaches were assessed using the Mann–Whitney–Wilcoxon test (Wilcoxon 1945; Mann & Whitney 1947). Genetic structure inferred by the different clustering approaches was compared using Pearson's correlation coefficient (*r*). Whenever applicable, statistical significance levels were adjusted for multiple comparisons using the Bonferroni procedure to correct for type I error (Weir 1996). These analyses were implemented in R (R Development Core Team 2013).

Results

Structure estimated by non-spatial approaches

Genetic structure inferred from the 711 Iberian honey bee individuals and the 309 neutral SNP data set using the Bayesian model-based clustering algorithm, implemented in STRUCTURE, and the model-free DAPC clustering algorithm, implemented in ADEGENET, is shown in Fig. VI - 2a (for $K = 2$) and Fig. Sup VI - 1 (Supporting information) (K = 3 to 5). The optimal number of clusters (K) varied between two, when estimated by ΔF_{st} and BIC, and four, when estimated by ΔK (Fig. Sup VI - 2, Supporting information). Incongruent optimal K values are often obtained by different methods (Campana *et al.* 2011), especially in the presence of low levels of population differentiation (Waples & Gaggiotti 2006), which is the case of the Iberian honey bee (global $\Phi_{PT} = 0.020$; pairwise Φ_{PT} values ranged from 0.000 to 0.046, but see Table Sup VI - 2 and Fig. Sup VI - 3, Supporting information for pairwise comparisons across the study area). Given that two of three measures agreed on an optimal $K = 2$ and the presence of two maternal lineages in Iberia (this and previous studies), it is likely that the number of clusters that best represents the maximum population structure is two.

At the optimal $K = 2$ (Fig. VI - 2a, Fig. Sup VI - 4, Supporting information), a concordant geographic pattern was produced by DAPC and STRUCTURE ($r = 0.79$, P-value = 0.0000 for individual Q values), although a deeper subdivision was inferred by the former than by the latter clustering approach, as measured by Q (P-value = 0.0000 for comparisons of individual Q values, Mann–Whitney–Wilcoxon test). Membership proportions estimated by STRUCTURE showed that most individuals (ranging from 53.3% in CT2 to 88.9% in MT1; see the bar below each clustering plot in Fig. VI - 2a) from sampling sites near the Pyrenees were assigned with high posterior probability ($Q \geq 0.80$) to the blue cluster. The percentages increased considerably when Q was inferred by DAPC, ranging from 50.0% in CT3 to 93.3% in MT2. Individuals with $Q \geq 0.80$ in the red cluster were common in the southern sampling sites of the Atlantic transect (ranging from 60.6% in AT5 to 63.3% in AT7) and rare in the Mediterranean transect (ranging from 0% in MT1-2 to 13.3% in MT6). However, again, the percentages increased considerably when Q was inferred by DAPC. Individuals exhibiting admixed proportions (Q to any cluster ≤ 0.80) prevailed in the northern part of the Atlantic transect and in the southern part of the central and Mediterranean transects when inferred by STRUCTURE, although they were rare when inferred by DAPC (Fig. VI - 2a).

When genetic structure was inferred from the complete SNP data set (Fig. VI - 2b), a deeper phylogeographical signal was captured by both clustering approaches, as measured by Q

(*P*-value = 0.0000 for comparisons of individual *Q* values inferred from the complete and neutral data sets with STRUCTURE and DAPC; Mann–Whitney–Wilcoxon test). Nonetheless, inclusion of the 74 putatively selected SNPs in the data set did not qualitatively change the overall geographic patterns of hybridization across Iberia, while providing additional support for an optimal $K = 2$, this time simultaneously estimated by the three methods ΔK, ΔF_{ST} and BIC (Fig. Sup VI - 2, Supporting information). The mean *Q* estimated with STRUCTURE from both the neutral and complete SNP data sets is shown at the sampling site level in Fig. VI - 2c (see Fig. Sup VI - 5, Supporting information for the corresponding DAPC plot). While further confirming the nearly concordant patterns across Iberia, this representation revealed a more abrupt transition from the blue to the red cluster in the central and Mediterranean transects than in the Atlantic transect, which suggests a contact zone located towards the northeastern part of Iberia.

Comparing maternal pattern with neutral structure

A median-joining network of a ~627-bp fragment of the $tRNA^{leu}$-cox2 intergenic mitochondrial region confirms the presence of the two highly divergent African (A) and western European (M) lineages in Iberia (Fig. 3a). The two lineages show a highly structured geographic pattern of distribution in Iberia. Mitotypes belonging to the M lineage were predominant in the northeastern half, whereas mitotypes belonging to the A lineage were fixed or almost fixed in the southwestern half of Iberia. While some sampling sites displayed a mixture of M and A mitotypes, the geographic distribution of the maternal lineages reveals a sharp northeastern–southwestern trend.

Membership proportions inferred by STRUCTURE (Fig. VI - 3b, c) and DAPC (Fig. Sup VI - 6, Supporting information) from neutral SNPs were contrasted with mtDNA data, at both sampling site and individual levels. At the sampling site level, the partitioning of neutral SNP variation into two clusters corresponded remarkably to M and A maternal lineages ($r = 0.81$ for both DAPC and STRUCTURE vs mtDNA; *P*-value < 0.0000). At the individual level, the correlations were weaker ($r = 0.46$ for DAPC, $r = 0.60$ for STRUCTURE), yet significant (*P*-value < 0.0000), suggesting differential gene flow among genomic compartments. Genome partitioning of individuals at greater values of *K* produced increasingly complex patterns (Fig. VI - 3c and Fig. Sup VI - 1, Supporting information). At $K = 4$, a pronounced east–west structuring of neutral variation was revealed. The Atlantic populations were clearly distinct from the other populations, and a north–south trend becomes more apparent in this transect. The nuclear pattern is consistent with maternal variation partitioned into African sublineages (Fig. VI - 3c). Sublineage AIII mitotypes were common in

northern Atlantic populations and were gradually replaced by sublineage AI mitotypes towards the south. In contrast to Atlantic populations, sublineage AIII mitotypes were virtually absent in populations of central and Mediterranean transects, which were dominated by sublineage AI mitotypes.

Structure estimated by spatial approaches

A number of studies have questioned the use of STRUCTURE for studying populations exhibiting continuous spatial distribution of genetic diversity (Serre & Pääbo 2004; Rosenberg *et al*. 2005). To address this issue, patterns of variation in Iberia were further investigated using spatially explicit approaches implemented by sPCA (Fig. VI - 4) and TESS (Fig. VI - 5). Analysis of neutral SNPs using sPCA showed that one global axis and one local axis were retained, indicating the existence of both global and local spatial structures in Iberia. The interpolation of the first global score, which was associated with a strong autocorrelation (Moran's I = 0.639), detected two clusters forming a cline (Fig. VI - 4a) concordant with nonspatial approaches. The second global score (Moran's I = 0.560) clearly differentiated the four northernmost sampling sites of the Atlantic transect and the southern half of central and Mediterranean transects (Fig. VI - 4b). The third global score (Moran's I = 0.443) further partitioned the Atlantic populations into two groups (north and south) and the southern half of central transect (Fig. VI - 4c). The northern half of the central transect was differentiated by the fourth global score producing a Moran's I = 0.392 (Fig. VI - 4d).

While the global test corroborated the presence of global spatial structure (max(t) = 0.0017; P-value = 0.0001), there was also structure at the local level (max(t) = 0.0019; P-value = 0.0001). The first local score (Moran's I = -0.075) highlighted the differences among individuals of northern Atlantic and central transects (Fig. VI - 4e), while the second local score (Moran's I = -0.071) differentiated the individuals from sampling sites in the middle part of the central transect (Fig. VI - 4f).

The additional spatial approach performed using TESS further confirmed the neutral patterns obtained previously (Fig. VI - 5). Two major clusters that largely overlapped those of nonspatial approaches (r = 0.76 for STRUCTURE vs TESS and r = 0.64 for DAPC vs TESS using individual Q values, P-value < 0.0000) were identified by TESS at each simulated K (Fig. Sup VI - 7, Supporting information). At the optimal K = 3 (Fig. Sup VI - 8, Supporting information), one additional minor cluster (mean Q = 0.023 in the white cluster) further partitioned the nuclear genomes of individuals mainly from the southern portion of the central transect (Fig. VI - 5a). While

TESS supported the major northeastern–southwestern cline and the contrasting patterns exhibited by Atlantic and Mediterranean populations, it did not capture the partitioning within the Atlantic transect, which was detected by the other clustering approaches and by mtDNA analysis.

The spatial patterns inferred from the complete SNP data set using both sPCA and TESS were largely concordant with those inferred from the neutral SNP data set, although, as observed with the nonspatial clustering approaches, a deeper phylogeographical signal was captured by the complete SNP data set (Figs Sup VI - 9 and Sup V - 10, Supporting information).

Geographic cline analysis

The geographic clines were modelled for 33 (17 neutral and 16 selected) individual SNPs, mean *Q* obtained with the complete and neutral SNP data sets, and mtDNA (Fig. VI - 6, Fig. Sup VI - 11, Supporting information). There was considerable variation in the identity of the best-fitting model among individual SNPs, SNP data sets and mtDNA (Table Sup VI - 3, Supporting information). The model '*Pmin/ Pmax* observed – no tails' was fitted to 16 of the 33 SNPs and to the neutral SNP data set, whereas '*Pmin/ Pmax* fixed – right tail' and '*Pmin/ Pmax* fixed – no tails' were fitted to the mtDNA and the complete SNP data set, respectively. The AICc values obtained for the null model of no clinal variation were higher for mtDNA (cline model = 255.1, null model = 516.3), complete SNP data set (cline model = 25.2, null model = 203.5) and neutral SNP data set (cline model = 33.2, null model = 137.2) than for any individual SNP (Table Sup VI - 3, Supporting information).

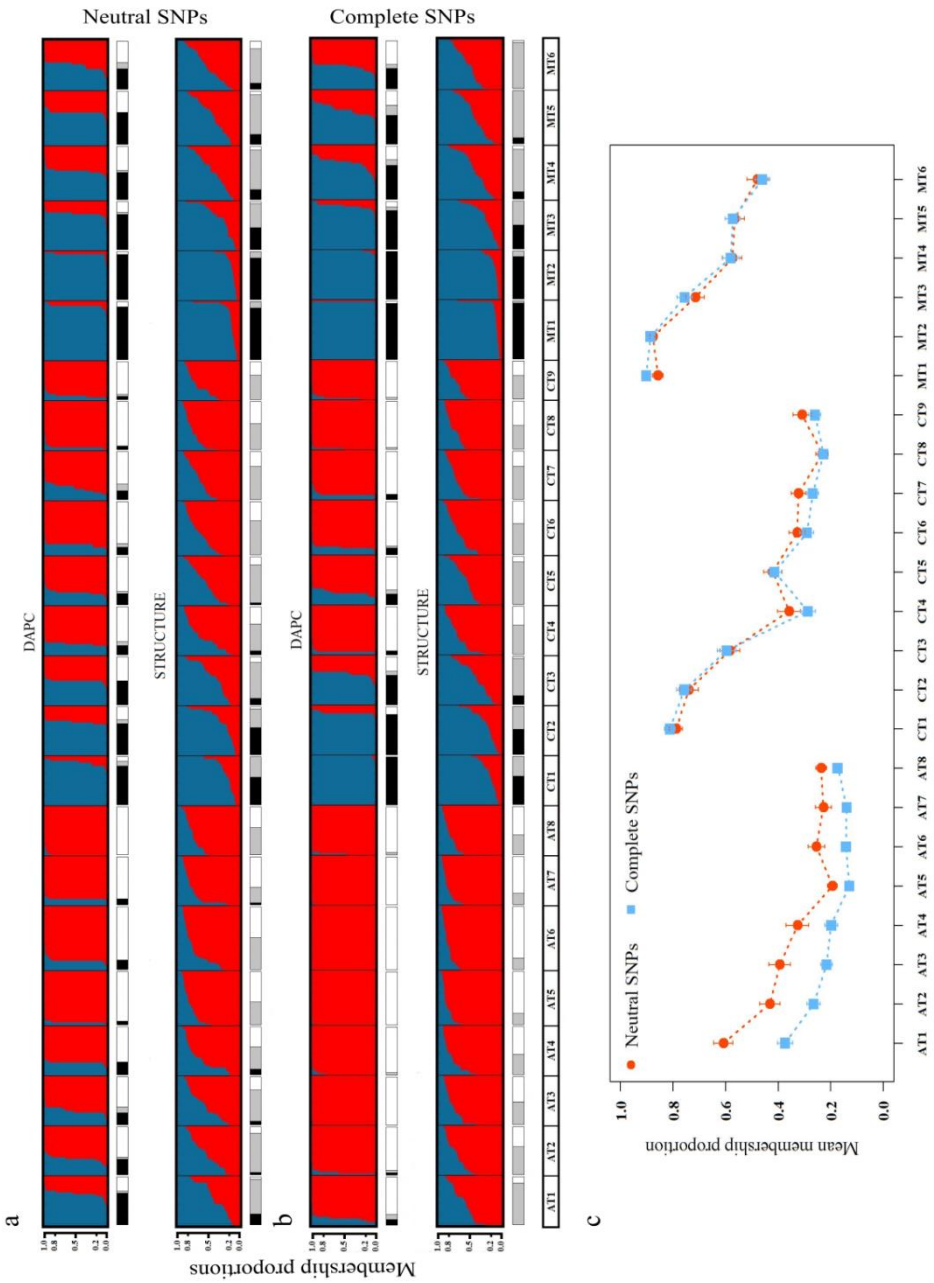


Figure VI - 2. Population structure of *A. m. iberiensis* estimated by DAPC and STRUCTURE at $K = 2$ clusters. The 23 sampling sites are arranged from north (AT1, CT1, MT1) to south (AT8, CT9, MT6) in each of the three transects (AT-Atlantic, CT-central, MT-Mediterranean). Plots represent each of the 711 individuals by a vertical bar partitioned into two colored segments (blue and red) corresponding to membership proportions (Q) in each of the two clusters. Black lines separate individuals from the 23 sampling sites, which are arranged from high Q (left) to low Q (right) in the blue cluster. The frequency of individuals per sampling site exhibiting $Q \geq 0.80$ (black), $0.20 > Q > 0.80$ (gray), and $Q \leq 0.20$ (white) in the blue cluster are indicated below each plot. Structure estimated from: (a) the neutral SNP dataset (309 loci) and (b) the complete SNP dataset (309 neutral plus 74 putatively selected loci identified by Chávez-Galarza *et al.* 2013). (c) Mean membership proportion (±SE) in the blue cluster inferred from the neutral and complete SNP datasets with STRUCTURE for each sampling site.

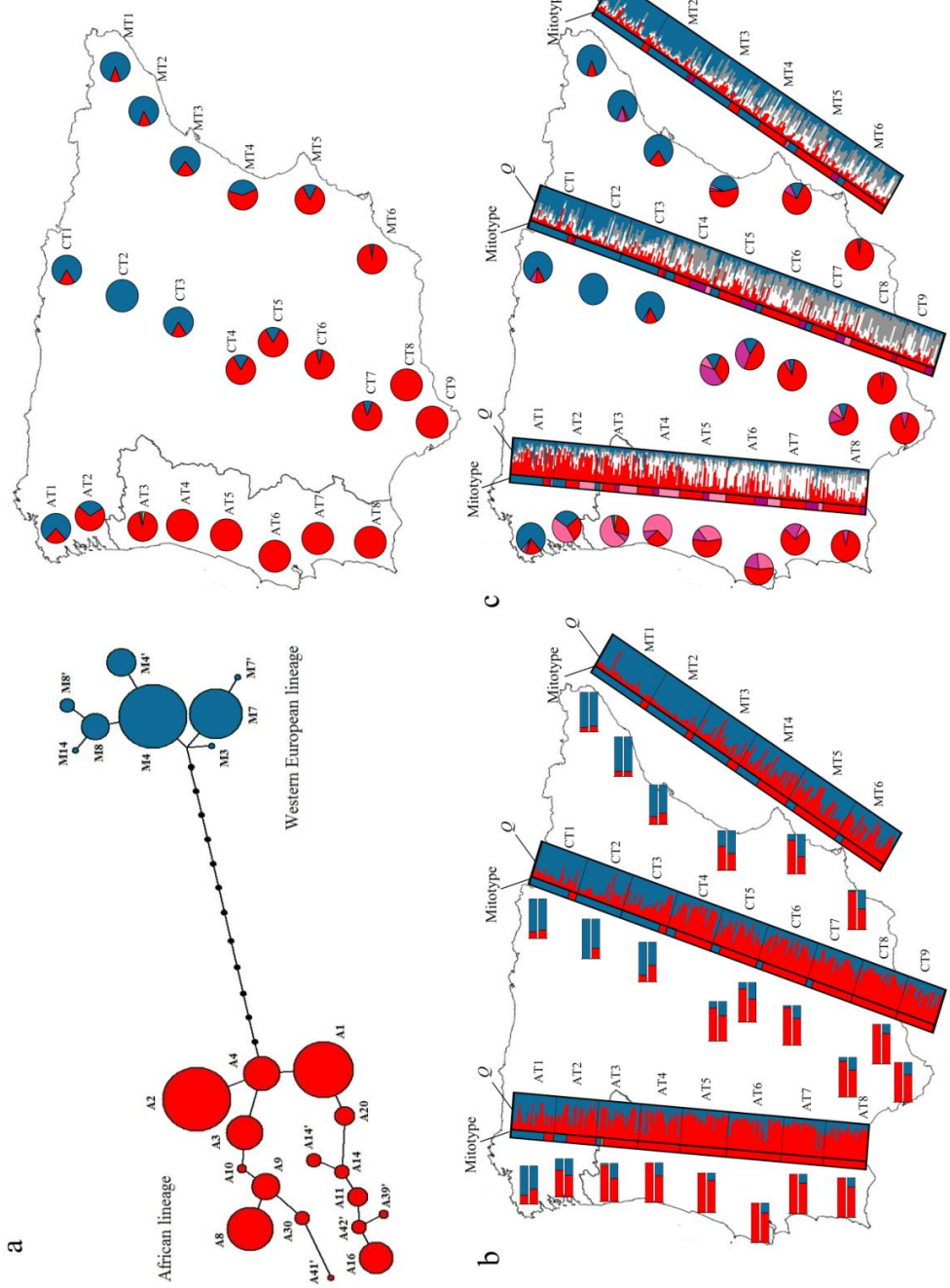


Figure VI - 3. Maternal pattern and estimated structure inferred from the neutral SNP dataset (309 loci) with STRUCTURE. Patterns of variation are displayed at both individual and sampling site level for each transect (AT-Atlantic, CT-central, MT-Mediterranean). (a) Median-joining network relating the mtDNA sequences of a ~627 bp fragment of the tRNA^{leu}-cox2 intergenic region. The African (A) and western European (M) mitotypes form two divergent lineages. The sizes of the circles size are proportional to the mitotype frequencies. The pie charts displayed in the Iberian map at the right side show the frequencies of the A and M mitotypes at each sampling site. (b) Patterns shown at $K = 2$ clusters. Vertical plots display the mitotype (A in red; M in blue; C in orange, one single individual in AT3) and the membership proportions (Q) for each of the 711 individuals. Horizontal bar plots show mitotype frequencies (top) and the mean Q in blue and red clusters (bottom) at each sampling site. (c) Patterns shown at $K = 4$ clusters. Vertical plots display the mitotype and Q for each of the 711 individuals. Maternal data is represented by M lineage (blue), C lineage (orange) and A sub-lineages (A_I in red, A_{II} in magenta, and A_{III} in pink). Pie charts show mitotype frequencies at each sampling site.

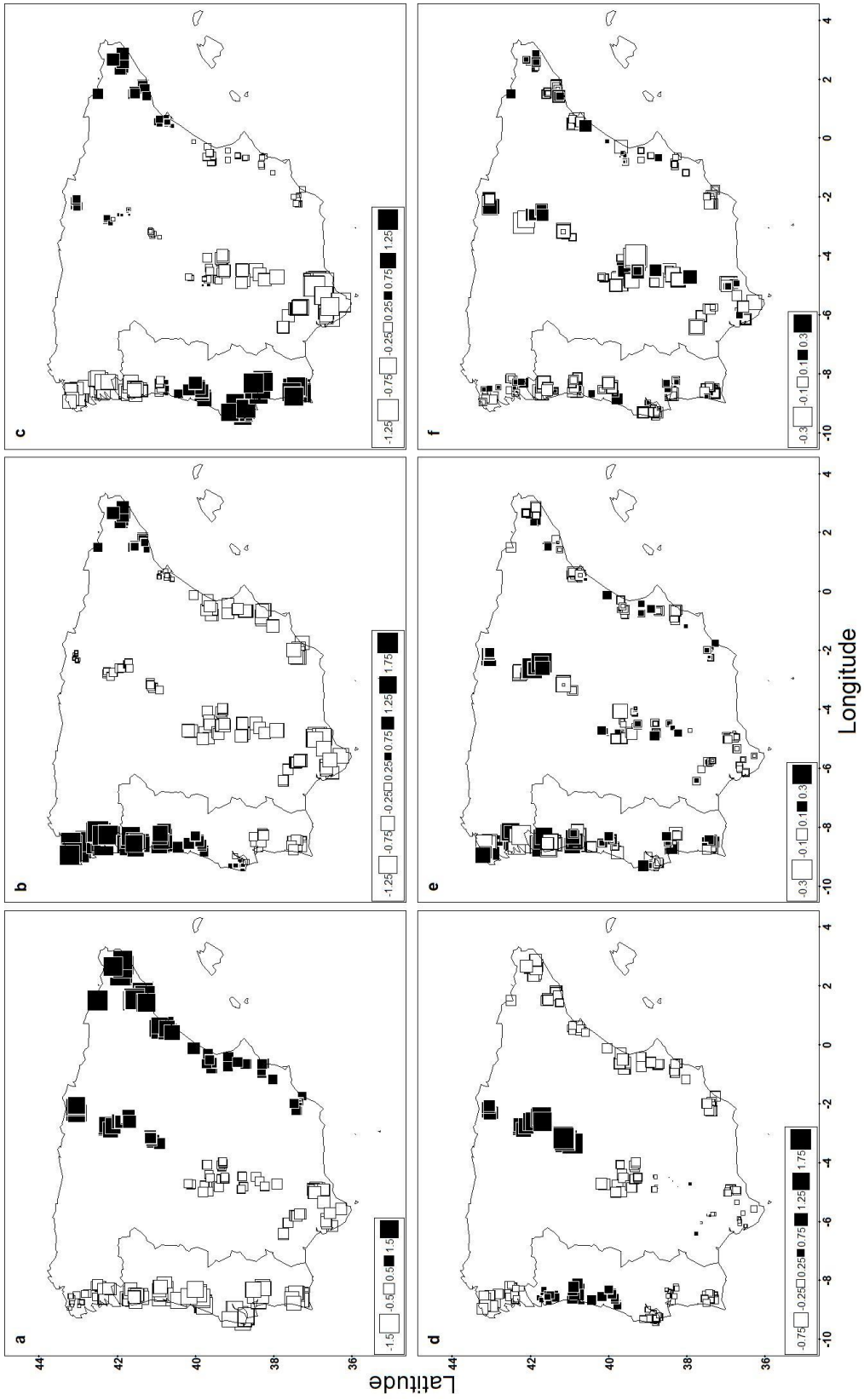


Figure VI - 4. Analysis of global and local structures, among 711 individuals of *A. m. iberiensis* from 23 sampling sites, by spatial principal component analysis (sPCA) using 309 neutral SNPs. Each square represents the score of an individual, which is positioned by its spatial coordinates. (a-d) The first four global scores of sPCA. (e-f) The first two local scores of sPCA.

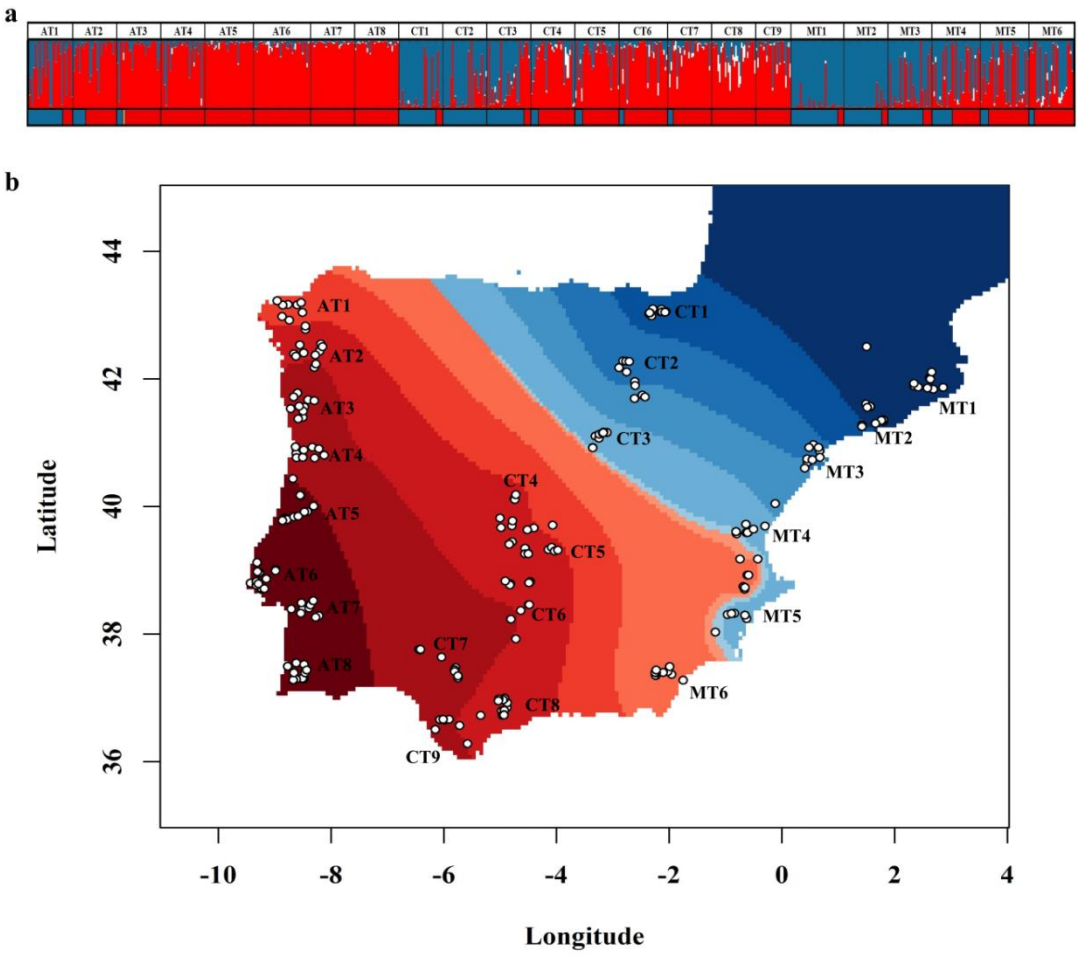


Figure VI - 5. Spatially explicit analysis implemented by the software TESS for the 711 individuals of *A. m. iberiensis* using the neutral SNP dataset (309 loci). (a) Plot of individuals' *Q* at the optimal K = 3 clusters. Each of the 711 individuals included in the analysis is represented by a vertical bar partitioned into three colored segments (blue, red, and white) corresponding to *Q* in each of the three clusters. Maternal data (M lineage in blue, C lineage in orange, and A lineage in red) are shown at the bottom. Sampling sites and individuals within sampling sites are arranged as in Fig. 3. (b) Map of the Iberian Peninsula showing the two major clusters ($Q \geq 0.5$) interpolated by TESS. Dots represent the locations of sampled apiaries across the Atlantic (AT1-8), central (CT1-9) and Mediterranean (MT1-6) transects.

Estimates of cline centre positions and widths for the 33 individual SNPs, complete SNP data set, neutral SNP data set and mtDNA were highly variable (Table Sup VI - 3, Supporting information). Coincidence analysis of the 33 SNPs revealed that 18 (Table Sup VI - 3, Supporting information), of which nine were neutral, could be constrained to share a common centre ($LRT_{same\text{-}diff.}$ = 26.88, 17 d.f., *P*-value > 0.05) at the consensus position of 665.2 km, as estimated by the likelihood profiles. The consensus centre of the 18 SNPs was coincident with those estimated for mtDNA (706.7 km) and for the complete (714.7 km) and neutral (725.7 km) SNP data sets ($LRT_{same\text{-}diff.}$ = 0–3.03, 1 d.f., P-value > 0.05 for all pairwise comparisons). Concordance analysis of the 33 SNPs revealed that 23 (Table Sup VI - 3, Supporting information), of which 11 were neutral, exhibited a similar width ($LRT_{same\text{-}diff.}$ = 1.60, 22 d.f., *P*-value > 0.05) of 1350 km. The consensus

width of the 23 SNPs was concordant ($LRT_{same\text{-}diff.}$ = 0.3–3.18, 1 d.f., *P*-value > 0.05 for all pairwise comparisons) with those estimated for the complete (1283.5 km) and neutral SNP data sets (1047.6 km), but not with that estimated for the mtDNA (580.9 km), which was significantly narrower ($LRT_{same\text{-}diff.}$ = 10.10–26.56, 1 d.f., *P*-value < 0.05 for all pairwise comparisons).

Linkage disequilibrium and genetic diversity

Genome-wide analysis of linkage disequilibrium (LD) between all possible pairs of neutral SNPs in each sampling site produced low levels of LD with mean r^2 values varying between 0.014 and 0.045 (Table Sup VI - 4, Supporting information). From a total of 701,362 pairwise comparisons, 14 687 pairs (2.09%), ranging from 1.49% to 3.41%, exhibited significant LD before Bonferroni correction (a single pair in CT4 remained significant after Bonferroni correction). Pairwise comparisons performed by linkage group also produced low LD values (data not shown). Levels of unbiased haploid diversity (*uh*) were low, ranging from 0.281 in the Atlantic transect (AT8) to 0.313 in the Mediterranean transect (M6, Table Sup VI - 4, Supporting information). Interestingly, despite the low levels of LD and uh across the study area, a trend of elevated values was observed towards the centre of the cline and overlapping the consensus centre location (Fig. VI - 7).

Discussion

Genetic studies of Iberian honey bees have revealed complex and often incongruent patterns of variation, which have led to competing hypothesis of primary intergradation (Ruttner *et al*. 1978) and secondary contact (Smith *et al*. 1991) as the leading mechanisms shaping patterns of variation. We examined a large number of individuals with a maternal locus and genome-wide SNPs and provided the most comprehensive portrait of clinal change across the entire Iberian honey bee distributional range. Our results support a signature of origin via secondary contact, which was still detectable despite intense beekeeping practices involving selective breeding and large-scale movement of colonies.

Nuclear and maternal patterns and cline origin

The multiple clustering approaches and the geographic cline analysis implemented on genome-wide SNPs collectively revealed a well-defined clinal pattern bisecting Iberia along a northeastern–southwestern axis, contrasting with the lack of microsatellite structure documented earlier (Franck *et al*. 1998; Cánovas *et al*. 2011; Miguel *et al*. 2011).

The most commonly suggested mechanisms underlying clinal patterns in gene frequencies are random genetic drift with isolation-by-distance effects, selection across an environmental gradient (primary intergradation), and secondary contact between previously isolated and genetically divergent populations. The SNP patterns exhibited by the Iberian honey bees could be explained by any of these mechanisms. However, several aspects of our data are more consistent with an alternative model of secondary contact and introgression between divergent populations previously isolated in glacial refugia, as proposed for a growing list of other Iberian taxa (reviewed by Gómez & Lunt 2007; Godinho *et al*. 2008; Gonçalves *et al*. 2009; Pinho *et al*. 2009; Miraldo *et al*. 2011; Carneiro *et al*. 2013; among others). First, while a deeper structure was retrieved by the complete SNP data set, excluding the 74 SNPs with signatures of selection (Chávez-Galarza *et al*. 2013) did not qualitatively change the clinal pattern of variation. A similar historical signal emerged when selected loci were removed from the genome-wide SNP data set.

Second, the geographic cline analysis revealed that a large proportion (9 of 17) of the neutral individual SNPs and both SNP data sets share a common cline centre, indicating considerable genome-wide coincidence. Existence of multiple coincident clines argues for secondary contact (Barton & Hewitt 1981), especially if some of the clines reflect changes in selectively neutral loci (Durrett *et al*. 2000). Simulations on the origin of contact zones show that a signature of secondary contact, which is characterized by clinal variation at neutral loci and extensive disequilibrium at the centre of the contact zone, can persist for thousands of generations if neutral loci were tightly linked to selected loci (Durrett *et al*. 2000). If neutral loci were not tightly linked to selected loci, the initially steep clines, formed at the moment the two divergent groups meet, will gradually widen as the intermixing proceeds. In contrast, in the presence of primary intergradation, neutral loci will not vary clinally and disequilibrium between a neutral locus and a closely linked locus under selection will decay quickly (Durrett *et al*. 2000).

Further support for secondary contact is provided by LD and diversity patterns. Both parameters show a trend of elevated values towards the centre of the cline and overlapping the consensus centre location, as expected when two divergent populations meet. It should be noted, however, that the LD levels at the centre of the contact zone were unexpectedly low for recent contact, which may simply reflect the sparse distribution of the SNPs and the short scale of LD in honey bees. Indeed, the exceptionally high recombination rate in honey bees (Beye *et al*. 2006) would lead to a rapid decay of LD after admixture, as observed in the Africanization process in the New World (Pinto *et al*. 2005).

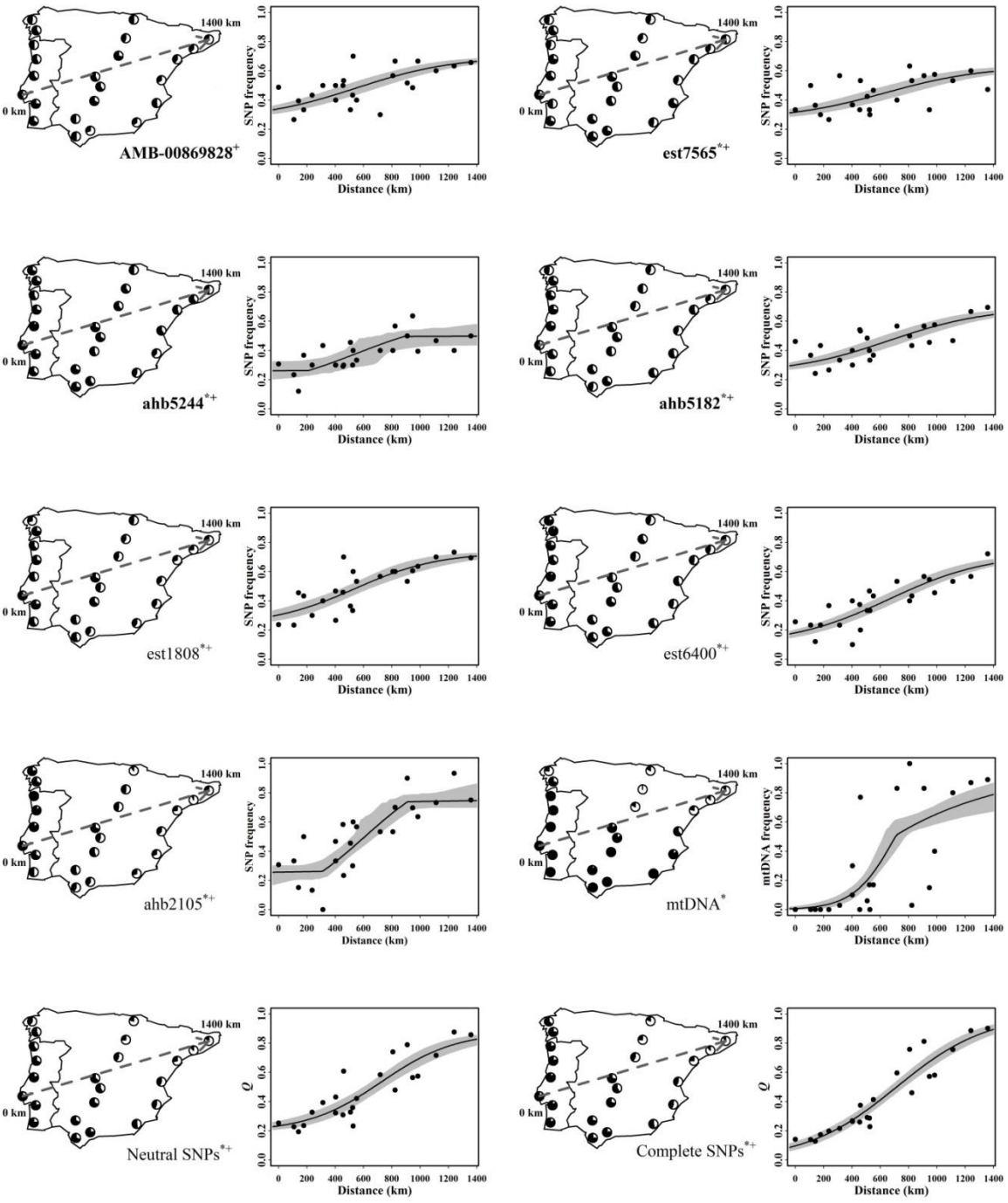


Figure VI - 6. Map of the Iberian Peninsula with pie charts summarizing frequency data for each sampling site and plot of maximum likelihood geographic cline for four neutral SNP loci (marked in bold), three selected SNP loci (see Fig. S V - 11 for the remaining 26 SNPs), mtDNA, and the *Q* values estimated with STRUCTURE from the neutral and the complete SNP datasets. The symbols · and · indicate the loci or datasets with coincident center and concordant width, respectively (see Table Sup VI - 3). The dashed line placed in each map represents the transect traced from Lisbon (0 km) to Girona (1400 km) for the geographic cline analysis.

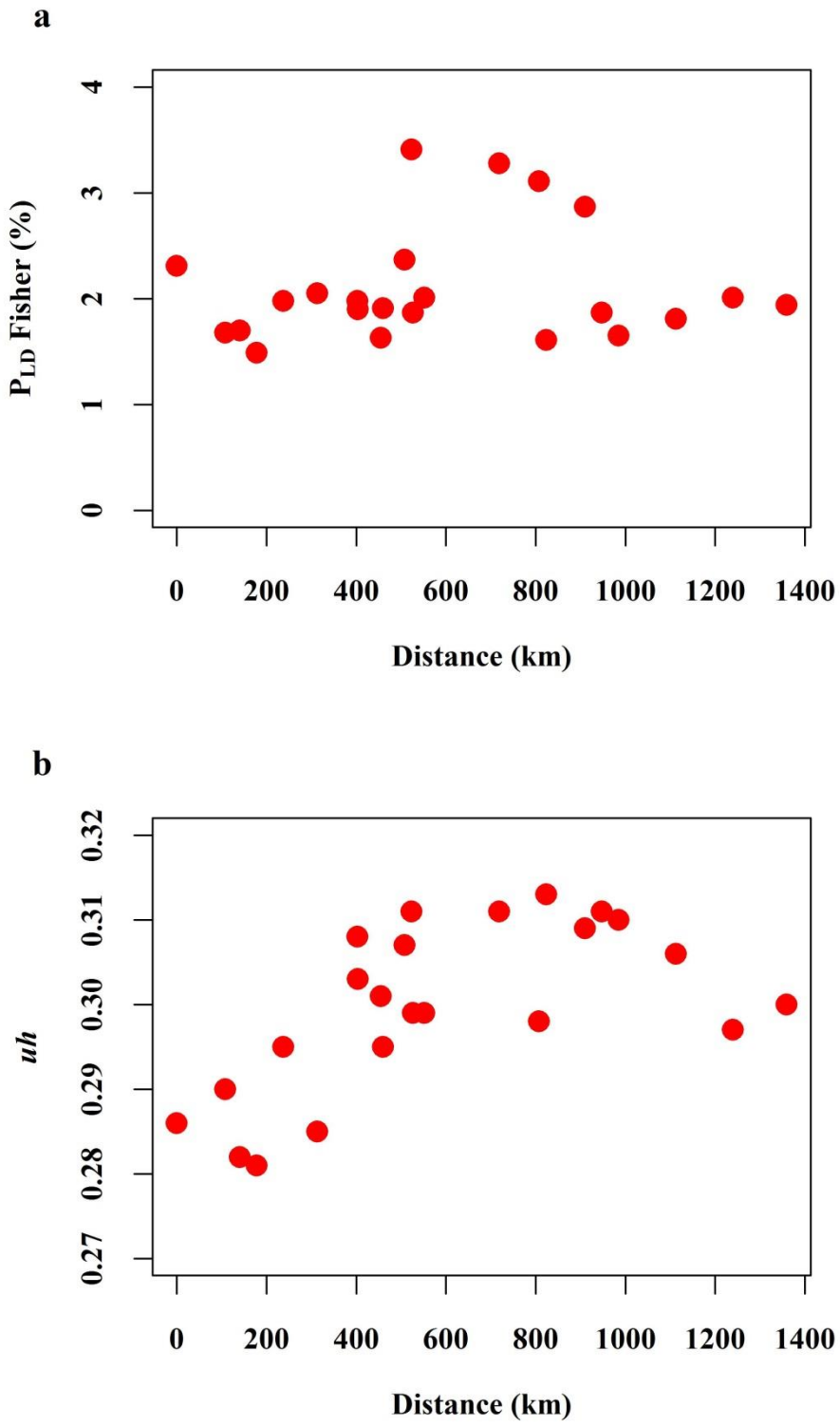


Figure VI - 7. (a) Percentage of pairs of neutral SNP loci showing significant linkage disequilibrium (LD) with the Fisher exact test (P_{LD} Fisher), before Bonferroni correction, and (b) unbiased haploid genetic diversity (*uh*) estimated from the neutral SNPs, both projected along the geographic cline.

Finally, while the observed patterns can also be explained by alternative processes involving isolation by distance with some migration or by divergence in parapatry, the ultimate support for secondary contact comes from mtDNA. Congruent with previous surveys (Smith *et al*. 1991; Garnery *et al*. 1992, 1995; Franck *et al*. 1998; Cánovas *et al*. 2008; Miguel *et al*. 2007), two highly divergent mtDNA lineages (African and western European) were identified and these lineages form a cline that closely parallels that of genome-wide SNPs, as revealed by both the clustering and the cline analyses. The mtDNA cline centre coincided with that of both the complete and the neutral SNP data sets and with most individual SNPs (18 of 33). In contrast, the mtDNA cline width was not concordant with any of the SNPs (either individual or data sets), which largely formed wider clines. Narrower maternal clines could arise from stronger drift on the smaller Ne of the haploid marker (Polechova & Barton 2011), from divergent selection on mtDNA types (Yuri *et al*. 2009), or from the greater gene flow expected of nuclear loci (Endler 1977). Our results of narrower maternal clines compared to nuclear clines add to a body of research showing discordant cytonuclear transmission across contact zones (Gowen et al. 2014 and many references therein).

Whether secondary contact resulted from ancient range expansions from North Africa following climate amelioration of the last postglacial period (De la Rúa *et al*. 2002; Pinto *et al*. 2013) or from recent introductions of the North African subspecies *A. m. intermissa* by the Arabs during Muslim occupation (Franck *et al*. 1998) is a matter of debate. A STRUCTURE analysis found no signs of *A. m. intermissa* genes, belonging to the African lineage, in Iberian populations, excepting for a residual component detected in sampling sites CT8 and CT9 nearby the Strait of Gibraltar (see Fig. Sup V - 3, Supporting information in Chávez-Galarza *et al*. 2013). This observation together with a report of deep differentiation between *A. m. iberiensis* and *A. m. intermissa* SNPs (Whitfield *et al*. 2006) means that a recent colonization event would have to be accompanied by a complete replacement of the nuclear, but not the mitochondrial, genomes of colonizers. This hypothesis assumes long-term male-biased gene flow, which would erode a signal of subdivision at the nuclear but not at the maternal level. The problem is that the latter scenario is not consistent with honey bee reproductive biology, as females have an important role in long-distance dispersal (Winston 1987). A much earlier event is therefore more likely to have been responsible for the patterns we see today.

The complexities and incongruences of Iberian honey bee patterns revealed by distinct genetic markers suggest an ancient history of allopatric divergence in Iberian refugia followed by postglacial range expansions and secondary contact. Iberia served as an important refuge during

the cold periods of the Pleistocene in Europe (reviewed by Weiss & Ferrand 2007). During this epoch, repeated cycles of contraction into and expansion out of multiple refugia shaped diversity patterns of great complexity in a variety of Iberian animal taxa (see Miraldo *et al*. 2011 and references therein), among which the honey bee is seemingly no exception. Estimates of genetic divergence from mtDNA (Arias & Sheppard 1996) and whole-genome nuclear DNA (Wallberg *et al*. 2014) suggest that the split among the four honey bee evolutionary lineages occurred between 670 000 and 300 000 years ago, respectively. Accordingly, colonization of Iberia across the Strait of Gibraltar (Ruttner *et al*. 1978; Whitfield *et al*. 2006; Han *et al*. 2012), from an origin in either Africa (Whitfield *et al*. 2006) or western Asia (Wallberg *et al*. 2014), likely occurred during Middle Pleistocene. Given dispersal abilities of the honey bees, it is plausible that they dispersed across the Iberian territory during interglacial periods and retreated to refugia during the glacial periods. Evidence from comparative phylogeography suggests that multiple refugia existed in Iberia ('refugia within refugia' paradigm of Gómez & Lunt 2007). Two such refugia, one in the Mediterranean coast of northeastern Spain, possibly close to the Ebro valley, and another in the Betic ranges of southern Spain, were inferred from overlapping subdivision patterns exhibited by several Iberian taxa (Gómez & Lunt 2007; see Fig. VI - 1). The blue and the red clusters identified in Fig. III are consistent with the existence of these two putative refugia. Although the presence of multiple honey bee refugia is a tentative result, it is an idea that can be further explored and tested using the power of multiple gene genealogies analysis.

Influence of human-mediated processes in shaping variation

Complicating the interpretation of diversity patterns in honey bees are contemporary human-mediated processes. Honey bees native to Europe have long been subjected to human manipulation (Crane 1999), with a variable impact in their genetic composition (reviewed by De la Rúa *et al*. 2009). In western Europe north of the Pyrenees, human-mediated movements of colonies between lineages (introduction of commercial queens) promoted variable levels of C-lineage introgression (Jensen *et al*. 2005; De la Rúa *et al*. 2009; Soland-Reckeweg *et al*. 2009; Oleksa *et al*. 2011; Pinto *et al*. 2014) and even replacement of the native *A. m. mellifera* subspecies in some areas (Jensen *et al*. 2005). In contrast, the Iberian honey bee is relatively free of C-lineage genes (Miguel *et al*. 2007, 2011; Cánovas *et al*. 2008, 2011; Pinto *et al*. 2013). A single colony harbouring a C-derived mitotype was scored in this study, and no signs of

introgression were detected at the nuclear level in that colony and in the remaining 710 (but see Figs. Sup V - 3 and Sup V - 4, Supporting information in Chávez-Galarza *et al*. 2013).

While movements of colonies between lineages have not yet seriously threatened the Iberian honey bee genetic integrity, the lack of microsatellite structure (Franck *et al*. 1998; Cánovas *et al*. 2011; Miguel *et al*. 2011) has been interpreted as an indication of high levels of gene flow aided by within-lineage movements associated with transhumance (Cánovas *et al*. 2011). This interpretation, however, is inconsistent with our results that show congruent cytonuclear subdivision, local structure detected by the sPCA, and relatively low levels of LD, none of which support the large-scale influence of transhumance in the Iberian honey bee gene pool. A possible explanation for microsatellite patterns is that saturation of the mutation spectrum homogenized allele size distributions (Nauta & Weissing 1996). Such homogenization has been suggested before to explain a similar pattern in the European rabbit (Queney *et al*. 2001). Almost identical to what is observed for the Iberian honey bee, the European rabbit exhibits a northeastern–southwestern cline for mtDNA (Branco *et al*. 2000), allozymes (Campos *et al*. 2007; Ferrand & Branco 2007), and nuclear sequence data (Branco *et al*. 2002; Geraldes *et al*. 2008; Carneiro *et al*. 2013), but no clinal pattern for microsatellites (Queney *et al*. 2001).

The cytonuclear structure in Iberian honey bees is noteworthy given that in Spain, over one million colonies, representing ~50% of existing colonies, have been yearly involved in wide-range movements, in the last decades (A. G. Pajuelo, personal communication). The fact that a marked clinal pattern in both the nuclear and mitochondrial genomes still persists indicates that human-mediated movements play a minor role in shaping Iberian honey bee genetic structure. Nonmutually exclusive explanations can be accounted for the observed pattern. Either transhumance takes place after the reproductive season, or some kind of reproductive barrier or local adaptation is preventing gene flow and long-term establishment of translocated colonies.

Concluding remarks

In this study, a well-defined northeastern–southwestern clinal pattern, revealed simultaneously by nuclear and maternal markers, provided support for the hypothesis of secondary contact proposed by earlier mtDNA studies. This finding, together with putative signatures of selection detected in a previous study (Chávez-Galarza *et al*. 2013), suggests a complex interplay between adaptation and demography in shaping the Iberian honey bee patterns that we see today. Contemporary human-mediated processes do not seem to be dramatically changing these patterns, a scenario that might

change if Spanish and Portuguese beekeepers adopt a strategy of using commercial C-lineage strains, as is occurring in several countries of western Europe (reviewed by De la Rúa *et al*. 2009; Pinto *et al*. 2014). Iberian honey bees are providers of important environmental services through pollination and are number one honey producers in the European Union (European Commission 2013). More importantly, Iberian honey bees represent an important reservoir of diversity that not long ago colonized a broad territory in western Europe (Franck *et al*. 1998; Garnery *et al*. 1998a; Miguel *et al*. 2007). Understanding patterns and underlying processes shaping Iberian honey bee's diversity is an important first step towards preserving this subspecies and thereby the species *Apis mellifera*, an effort of unquestionable value as we face a worldwide honey bee crisis.

References

Arias MC, Sheppard WS (1996) Molecular phylogenetics of honey bee subspecies (*Apis mellifera* L.) inferred from mitochondrial DNA sequence. *Molecular Phylogenetics and Evolution*, **5**, 557–566.

Arias MC, Rinderer TE, Sheppard WS (2006) Further characterization of honey bees from the Iberian Peninsula by allozyme, morphometric and mtDNA haplotype analyses. *Journal of Apicultural Research*, **45**, 188–196.

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37–48.

Barton NH, Gale KS (1993) Genetic analysis of hybrid zones. In: Hybrid Zones and the Evolutionary Process (ed. Harrison RG), pp. 13–42. Oxford University Press, Oxford.

Barton NH, Hewitt GM (1981) The genetic basis of hybrid inviability in the grasshopper *Podisma pedestris*. *Heredity*, **47**, 367–383.

Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, **16**, 113–148.

Barton NH, Hewitt GM (1989) Adaptation, speciation and hybrid zones. *Nature*, **341**, 497–503.

Beye M, Gattermeier I, Hasselmann M *et al*. (2006) Exceptionally high levels of recombination across the honey bee genome. *Genome Research*, **16**, 1339–1344.

Branco M, Ferrand N, Monnerot M (2000) Phylogeography of the European rabbit (*Oryctolagus cuniculus*) in the Iberian Peninsula inferred from RFLP analysis of the cytochrome b gene. *Heredity*, **85**, 307–317.

Branco M, Monnerot M, Ferrand N, Templeton AR (2002) Postglacial dispersal of the European rabbit (*Oryctolagus cuniculus*) on the Iberian peninsula reconstructed from nested clade and mismatch analyses of mitochondrial DNA genetic variation. *Evolution*, **56**, 792–803.

Campana MG, Hunt HV, Jones H, White J (2011) CorrSieve: software for summarizing and evaluating Structure output. *Molecular Ecology Resources*, **11**, 349–352.

Campos R, Branco M, Weiss S, Ferrand N (2007) Patterns of hemoglobin polymorphism (α-globin (HBA) and β-globin (HBB)) across the contact zone of two distinct phylogeographic lineages of the European rabbit (*Oryctolagus cuniculus*). In: Phylogeography of Southern European Refugia (eds Weiss S, Ferrand N), pp. 237–255. Springer, The Netherlands.

Cánovas F, De la Rúa P, Serrano J, Galián J (2008) Geographical patterns of mitochondrial DNA variation in *Apis mellifera iberiensis* (Hymenoptera: Apidae). *Journal of Zoological Systematics and Evolutionary Research*, **46**, 24–30.

Cánovas F, De la Rúa P, Serrano J, Galián J (2011) Microsatellite variability reveals beekeeping influences on Iberian honeybee populations. *Apidologie*, **42**, 235–251.

Carneiro M, Stuart JEB, Afonso S *et al.* (2013) Steep clines within a highly permeable genome across a hybrid zone between two subspecies of the European rabbit. *Molecular Ecology*, **22**, 2511–2525.

Carranza S, Arnold EN, Wade E, Fahd S (2004) Phylogeography of the false smooth snakes, *Macroprotodon* (Serpentes, Colubridae): mitochondrial DNA sequences show European populations arrived recently from Northwest Africa. *Molecular Phylogenetics and Evolution*, **33**, 523–532.

Chávez-Galarza J, Henriques D, Johnston JS *et al.* (2013) Signatures of selection in the Iberian honey bee (*Apis mellifera iberiensis*) revealed by a genome scan analysis of single nucleotide polymorphisms. *Molecular Ecology*, **22**, 5890–5907.

Chen C, Durand E, Forbes F, François O (2007) Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Molecular Ecology Notes*, **7**, 747–756.

Cornuet JM, Fresnaye J (1989) Biometrical study of honey bee populations from Spain and Portugal. *Apidologie*, **20**, 93–101.

Cosson JF, Hutterer R, Libois R *et al.* (2005) Phylogeographical footprints of the Strait of Gibraltar and Quaternary climatic fluctuations in the western Mediterranean: a case study with the

greater white-toothed shrew, *Crocidura russula* (Mammalia: Soricidae). *Molecular Ecology*, **14**, 1151–1162.

Crane E (1999) The World History of Beekeeping and Honey Hunting. Routledge, New York.

De la Rúa P, Galián J, Serrano J, Moritz RFA (2001) Genetic structure and distinctness of *Apis mellifera* L. populations from the Canary Islands. *Molecular Ecology*, **10**, 1733–1742.

De la Rúa P, Galián J, Serrano J, Moritz RFA (2002) Microsatellite analysis of non-migratory colonies of *Apis mellifera iberica* from south-eastern Spain. *Journal of Zoological Systematics and Evolutionary Research*, **40**, 164–168.

De la Rúa P, Galián J, Serrano J, Moritz RFA (2003) Genetic structure of Balearic honeybee populations based on microsatellite polymorphism. *Genetics Selection Evolution*, **35**, 339–350.

De la Rúa P, Hernandez-Garcia R, Jiménez Y, Galián J, Serrano J (2005) Biodiversity of *Apis mellifera iberica* (Hymenoptera: Apidae) from northeastern Spain assessed by mitochondrial analysis. *Insect Systematics & Evolution*, **36**, 21–28.

De la Rúa P, Jaffé R, Dall'Olio R, Muñoz I, Serrano J (2009) Biodiversity, conservation and current threats to European honeybees. *Apidologie*, **40**, 263–284.

De la Rúa P, Jiménez Y, Galián J, Serrano J (2004) Evaluation of the biodiversity of honey bee (*Apis mellifera*) populations from eastern Spain. *Journal of Apicultural Research*, **43**, 162–166.

Derryberry EP, Derryberry G, Maley J, Brumfield RT (2014) HZAR: hybrid zone analysis using an R software package. *Molecular Ecology Resources*, **14**, 652–663.

Durrett R, Buttel L, Harrison R (2000) Spatial models for hybrid zones. *Heredity*, **84**, 9–19.

Earl DA, Von Holdt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Endler JA (1977) Geographic Variation, Speciation, and Clines. Princeton University Press, Princeton, NJ.

Engel MS (1999) The taxonomy of recent and fossil honey bees (Hymenoptera : Apidae; *Apis*). *Journal of Hymenoptera Research*, **8**, 165–196.

European Commission (2013) EU Market Situation for Honey (ed. Agriculture and Rural Development), pp. 1–20.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Ferrand N, Branco M (2007) The evolutionary history of the European rabbit (*Oryctolagus cuniculus*): major patterns of population differentiation and geographic expansion inferred from protein polymorphism. In: Phylogeography of Southern European Refugia (eds Weiss S, Ferrand N), pp. 207–235. Springer, The Netherlands.

Franck P, Garnery L, Celebrano G, Solignac M, Cornuet JM (2000) Hybrid origins of honey bees from Italy (*Apis mellifera ligustica*) and sicily (*A. m. sicula*). *Molecular Ecology*, **9**, 907–921.

Franck P, Garnery L, Loiseau A *et al*. (2001) Genetic diversity of the honey bee in Africa: microsatellite and mitochondrial data. *Heredity*, **86**, 420–430.

Franck P, Garnery L, Solignac M, Cornuet JM (1998) The origin of west European subspecies of honeybees (*Apis mellifera*): new insights from microsatellite and mitochondrial data. *Evolution*, **52**, 1119–1134.

Garnery L, Cornuet JM, Solignac M (1992) Evolutionary history of the honey bee *Apis mellifera* inferred from mitochondrial DNA analysis. *Molecular Ecology*, **1**, 145–154.

Garnery L, Franck P, Baudry E *et al*. (1998a) Genetic diversity of the west European honey bee (*Apis mellifera mellifera* and *A. m. iberica*). I. Mitochondrial DNA. *Genetics Selection Evolution*, **30**, S31–S47.

Garnery L, Franck P, Baudry E *et al*. (1998b) Genetic diversity of the west European honey bee (*Apis mellifera mellifera* and *A. m. iberica*). II. Microsatellite loci. *Genetics Selection Evolution*, **30**, S49–S74.

Garnery L, Mosshine EH, Oldroyd BP, Cornuet JM (1995) Mitochondrial DNA variation in Moroccan and Spanish honey bee populations. *Molecular Ecology*, **4**, 465–471.

Garnery L, Solignac M, Celebrano G, Cornuet JM (1993) A simple test using restricted PCR-amplified mitochondrial DNA to study the genetic structure of *Apis mellifera* L. *Experientia*, **49**, 1016–1021.

Gay L, Crochet P-A, Bell DA, Lenormand T (2008) Comparing clines on molecular and phenotypic traits in hybrid zones: a window on tension zone models. *Evolution*, **62**, 2789–2806.

Geraldes A, Carneiro M, Delibes-Mateos M, Villafuerte R, Nachman MW, Ferrand N (2008) Reduced introgression of the Y chromosome between subspecies of the European rabbit (*Oryctolagus cuniculus*) in the Iberian Peninsula. *Molecular Ecology*, **17**, 4489–4499.

Godinho R, Crespo EG, Ferrand N (2008) The limits of mtDNA phylogeography: complex patterns of population history in a highly structured Iberian lizard are only revealed by the use of nuclear markers. *Molecular Ecology*, **17**, 4670–4683.

Gómez A, Lunt DH (2007) Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. In: Phylogeography of Southern European Refugia (eds Weiss S, Ferrand N), pp. 155–188. Springer, The Netherlands.

Gonçalves H, Marínez-Solano I, Pereira RJ, Carvalho B, García-París M, Ferrand N (2009) High levels of population subdivision in a morphologically conserved Mediterranean toad (*Alytes cisternasii*) result from recent, multiple refugia: evidence from mtDNA, microsatellites and nuclear genealogies. *Molecular Ecology*, **18**, 5143–5160.

Gowen FC, Maley JM, Cicero C *et al*. (2014) Speciation in Western Scrub-Jays, Haldane's rule, and genetic clines in secondary contact. *BMC Evolutionary Biology*, **14**, 135.

Guillaumet A, Pons JM, Godelle B, Crochet PA (2006) History of the Crested Lark in the Mediterranean region as revealed by mtDNA sequences and morphology. *Molecular Phylogenetics and Evolution*, **39**, 645–656.

Han F, Wallberg A, Webster MT (2012) From where did the Western honeybee (*Apis mellifera*) originate? *Ecology and Evolution*, **2**, 1949–1957.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. Nature, 405, 907–913.

Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. TAG. *Theoretical and Applied Genetics*, **38**, 226–231.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.

Jensen AB, Palmer KA, Boomsma JJ, Pedersen BV (2005) Varying degrees of *Apis mellifera ligustica* introgression in protected populations of the black honeybee, *Apis mellifera mellifera*, in northwest Europe. *Molecular Ecology*, **14**, 93–106.

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, **11**, 94.

Jombart T, Devillard S, Dufour AB, Pontier D (2008) Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity*, **101**, 92–103.

Kandemir I, Kence M, Sheppard WS, Kence A (2006) Mitochondrial DNA variation in honey bee (*Apis mellifera* L.) populations from Turkey. *Journal of Apicultural Research*, **45**, 33–38.

Klaassen CH, Gibbons JG, Fedorova ND, Meis JF, Rokas A (2012) Evidence for genetic differentiation and variable recombination rates among Dutch populations of the opportunistic human pathogen *Aspergillus fumigatus*. *Molecular Ecology*, **21**, 57–70.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, **4**, 981–994.

Mann HB, Whitney DR (1947) On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, **18**, 59–60.

Meixner MD, Leta MA, Koeniger N, Fuchs S (2011) The honey bees of Ethiopia represent a new subspecies of *Apis mellifera-Apis mellifera simensis* n. ssp. *Apidologie*, **42**, 425–437.

Miguel I, Baylac M, Iriondo M *et al*. (2011) Both geometric morphometric and microsatellite data consistently support the differentiation of the *Apis mellifera* M evolutionary branch. *Apidologie*, **42**, 150–161.

Miguel I, Iriondo M, Garnery L, Sheppard WS, Estonba A (2007) Gene flow within the M evolutionary lineage of *Apis mellifera*: role of the Pyrenees, isolation by distance and postglacial re-colonization routes in the western Europe. *Apidologie*, **38**, 141–155.

Miraldo A, Hewitt GM, Paulo OS, Emerson BC (2011) Phylogeography and demographic history of *Lacerta lepida* in the Iberian Peninsula: multiple refugia, range expansions and secondary contact zones. *BMC Evolutionary Biology*, **11**, 170.

Nauta MJ, Weissing FJ (1996) Constraints on allele size at microsatellite loci: implications for genetic differentiation. *Genetics*, **143**, 1021–1032.

Nielsen D, Page RE Jr, Crosland MW (1994) Clinal variation and selection of MDH allozymes in honey bee populations. *Experientia*, **50**, 867–871.

Oleksa A, Chybicki I, Tofilski A, Burczyk J (2011) Nuclear and mitochondrial patterns of introgression into native dark bees (*Apis mellifera mellifera*) in Poland. *Journal of Apicultural Research*, **50**, 116–129.

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics*, **28**, 2537–2539.

Phillips BL, Baird SJ, Moritz C (2004) When vicars meet: a narrow contact zone between morphologically cryptic phylogeographic lineages of the rainforest skink, *Carlia rubrigularis*. *Evolution*, **58**, 1536–1548.

Pinho C, Kaliontzopoulou A, Carretero MA, Harris DJ, Ferrand N (2009) Genetic admixture between the Iberian endemic lizards *Podarcis bocagei* and *Podarcis carbonelli*: evidence for limited natural hybridization and a bimodal hybrid zone. *Journal of Zoological Systematics and Evolutionary Research*, **47**, 368–377.

Pinto MA, Henriques D, Chávez-Galarza J *et al*. (2014) Genetic integrity of the Dark European honey bee (*Apis mellifera mellifera*) from protected populations: a genome-wide assessment using SNPs and mtDNA sequence data. *Journal of Apicultural Research*, **53**, 269–278.

Pinto MA, Henriques D, Neto M *et al*. (2013) Maternal diversity patterns of Ibero-Atlantic populations reveal further complexity of Iberian honey bees. *Apidologie*, **44**, 430–439.

Pinto MA, Muñoz I, Chávez-Galarza J, De la Rúa P (2012) The Atlantic side of the Iberian Peninsula: a hot-spot of novel African honey bee maternal diversity. *Apidologie*, **43**, 663–673.

Pinto MA, Rubink WL, Patton JC, Coulson RN, Johnston JS (2005) Africanization in the United States: replacement of Feral European Honeybees (*Apis mellifera* L.) by an African Hybrid Swarm. *Genetics*, **170**, 1653–1665.

Polechova J, Barton N (2011) Genetic drift widens the expected cline but narrows the expected cline width. *Genetics*, **189**, 227–235.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Queney G, Ferrand N, Weiss S, Mougel F, Monnerot M (2001) Stationary distributions of microsatellite loci between divergent population groups of the European rabbit (*Oryctolagus cuniculus*). *Molecular Biology and Evolution*, **18**, 2169–2178.

R Development Core Team (2013) R: A language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rortais A, Arnold G, Alburaki M, Legout H, Garnery L (2011) Review of the DraI COI-COII test for the conservation of the black honeybee (*Apis mellifera mellifera*). *Conservation Genetics Resources*, **3**, 383–391.

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.

Rosenberg NA, Mahajan S, Ramachandran S *et al*. (2005) Clines, clusters, and the effect of study design on the inference of human population structure. *PLoS Genetics*, **1**, e70.

Ruttner F (1988) Biogeography and Taxonomy of Honey Bees. Springer, Berlin.

Ruttner F, Tassencourt L, Louveaux J (1978) Biometrical-statistical analysis of the geographic variability of *Apis mellifera* L. *Apidologie*, **9**, 363–381.

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular Cloning, 2nd edn. Cold Spring Harbor Laboratory Press, New York.

Serre D, Pääbo S (2004) Evidence for gradients of human genetic diversity within and among continents. *Genome Research*, **14**, 1679–1685.

Shaibi T, Muñoz I, Dall'Olio R *et al*. (2009) *Apis mellifera* evolutionary lineages in Northern Africa: Libya, where orient meets occident. *Insectes Sociaux*, **56**, 293–300.

Sheppard WS, Meixner MD (2003) *Apis mellifera pomonella*, a new honey bee subspecies from Central Asia. *Apidologie*, **34**, 367–375.

Smith DR, Glenn TC (1995) Allozyme polymorphisms in Spanish honeybees (*Apis mellifera iberica*). *Journal of Heredity*, **86**, 12–16.

Smith DR, Palopoli MF, Taylor BR *et al*. (1991) Geographical overlap of two mitochondrial genomes in Spanish honeybees (*Apis mellifera iberica*). *Journal of Heredity*, **82**, 96–100.

Soland-Reckeweg G, Heckel G, Neumann P, Fluri P, Excoffier L (2009) Gene flow in admixed populations and implications for the conservation of the Western honeybee, *Apis mellifera*. *Journal of Insect Conservation*, **13**, 317–328.

Szymura JM, Barton NH (1986) Genetic analysis of a hybrid zone between the fire-bellied toads, *Bombina bombina* and *B. variegata*, near Cracow in southern Poland. *Evolution*, **40**, 1141–1159.

Szymura JM, Barton NH (1991) The genetic structure of the hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*: comparisons between transects and between loci. *Evolution*, **45**, 237–261.

Tamura K, Peterson D, Peterson N *et al*. (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Yuri T, Jernigan RW, Brumfield RT, Bhagabati NK, Braun MJ (2009) The effect of marker choice on estimated levels of introgression across an avian (Pipridae: *Manacus*) hybrid zone. *Molecular Ecology*, **18**, 4888–4903.

Wallberg A, Han F, Wellhagen G *et al*. (2014) A worldwide survey of genome sequence variation provides insight into the evolutionary history of the honeybee *Apis mellifera*. *Nature Genetics*, **46**, 1081–1088.

Waples RS, Gaggiotti O (2006) Invited review: What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, **15**, 1419–1439.

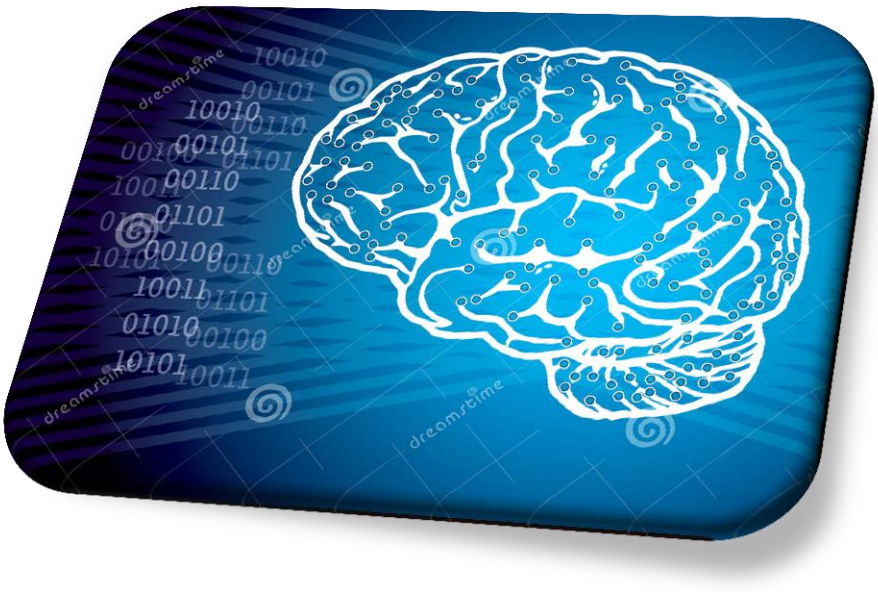
Weir B (1996) Genetic Data Analysis. Sinauer Associates, Sunderland.

Weiss S, Ferrand N (2007) Current perspectives in phylogeography and the significance of South European refugia in the creation and maintenance of European biodiversity. In: Phylogeographyof Southern European Refugia (eds Weiss S, Ferrand N), pp. 237–255. Springer, The Netherlands.

Whitfield CW, Behura SK, Berlocher SH *et al*. (2006) Thrice out of Africa: ancient and recent expansions of the honey bee, *Apis mellifera*. *Science*, **314**, 642–645.

Wilcoxon F (1945) Individual comparisons by ranking methods. *Biometrics Bulletin*, **1**, 80–83.

Winston M (1987) The biology of the Honey Bee. Harvard University Press, London.



Chapter VII

General conclusions

Concluding Remarks

Hybrid zones have been described as natural laboratories for evolutionary studies. In that sense, the Iberian Peninsula has functioned as a hybrid zone where genetically distinct populations meet, mate, and produce at least some offspring of mixed ancestry. At the same time, Iberia is considered a glacial refuge for many species during cold periods of the Quaternary, accomplishing an important role in their evolutionary history. Hybrid zone formation occurs either via primary intergradation along an environmental gradient or as a result of secondary contact between two populations that have previously differentiated in allopatry. Unfortunately, both processes can produce similar patterns of phenotypic and genetic variation, making it difficult or impossible in many cases to distinguish between these alternative hypotheses. As for the case of the Iberian honey bee, studies based on morphology, allozymes, pheromones and SNPs have proposed a process of primary intergradation, but mitochondrial analysis has suggested a process of secondary contact. Therefore, the goal of this study was to disentangle the complex variation patterns of the Iberian honey bee hybrid zone.

In this study, the *Dra*I test was used to analyze the haplotype pattern of Portuguese honey bee populations. To that end, 950 colonies, representing continental Portugal and the archipelagos of Azores and Madeira, were screened for mtDNA variation and novel band patterns were sequenced. The results revealed 16 novel haplotypes of African ancestry, 15 belonging to sub-lineage A_{III} and 1 to sub-lineage A_{I}. Such an impressive number of novel haplotypes, mostly of sub-lineage A_{III}, reveals that previous studies have missed an important diversity component in Iberian honey bees, suggesting that the Atlantic side of the Iberian Peninsula is an important reservoir of maternal diversity.

Building from the previous results, a fine-scale analysis was performed using $tRNA^{leu}$-cox2 sequence data for inferring maternal diversity and population structure across the entire Iberian Peninsula. To that end, 711 drones collected along three north-south transects in the Iberian Peninsula were analyzed. Sequence data of the $tRNA^{leu}$-cox2 intergenic region allowed identification of 164 novel haplotypes of lineages A and M origin and revealed a more complex diversity pattern in the Iberian honey bee than previously thought when using only the *Dra*I test the. At the same time, the sequencing of this region was crucial to demonstrate and to identify the underestimation of some haplotypes as determined by the *Dra*I test, and also its usefulness to trace the origin of haplotypes. The distribution of haplotypes A and M reported in this study has further refined the

Southwestern-Northeastern clinal pattern previously described and has also reinforced the hypothesis of a hybrid origin for the Iberian honey bee resulting from a process of secondary contact. The distribution of African sub-lineages was well characterized with (1) sub-lineage A_{I} more widely dispersed and highly frequent in Southern Iberia, (2) sub-lineage A_{II} mostly confined to the central part of Spain, and (3) sub-lineage A_{III} showing an Atlantic distribution. The complex network of haplotypes A and M, as well as the diversity pattern, suggest that M lineage was more affected by climatic oscillations during Pleistocene glacial periods than lineage A. Both lineages A and M would have mainly survived in two refuges located in the Northeastern (between the Iberian Mountain Range and Pyrenees) and Southern (Betic Ranges in Spain) of Iberia, as suggested for other taxa. Nonetheless, the confined distribution of sub-lineage A_{III} to the North Atlantic side of Iberia suggests existence of a putative third refuge. A highlighting finding was related with identification of haplotypes with intermediate A and M characteristics, which were denominated 'ancestral' due to their stronger phylogenetic relationship with lineage M. Furthermore, the phylogenetic topology suggests that sub-lineages A_{I}, A_{II}, A_{III} and lineage M have diverged of a common ancestral similar to sub-lineage Z. This finding suggests that lineage M would have a more ancestral African origin, not from North African populations but from Northeastern African and Near East populations where haplotypes of sub-lineage Z origin have been reported. The low number of colonies with haplotypes C (1 colony) suggests a low level of importation of commercial queens, which contrasts with northern Europe.

Some studies have suggested that the maternal distribution of lineages A and M has been shaped by natural selection. Therefore, a search for footprints of selection was carried out across the Iberian honey bee genome using SNPs. The 711 individuals analyzed for the $tRNA^{leu}$-cox2 intergenic region were screened for 1536 SNPs, of which only 383 were polymorphic. In this study, 69 SNP outliers were identified using two coalescent and two Bayesian methods but only 17 outliers were cross-validated by the four methods. Among the 17 outliers, 10 exhibited the strongest signal of selection (9 directional and 1 balancing). A spatial method was also used to identify outliers associated with environmental variables. This method identified 33 outliers, of which 28 were coincident with those found by the previous methods. Genomic information showed that some outliers were located in or near genes involved in a diverse array of putative functions including signaling, structural, metabolism, regulation, transport, and immunity. An important result is that some outliers detected were mapped to genes involved in the same biological function such as vision, xenobiotic detoxification and immune response. For the vision mechanism, five

outlier SNPs were detected, of which two (ahb8266 and ahb142) marked genes with structural function responsible for neural development, synapse formation and axon guidance in the process of vision, and the other three (est7297, est2423, ahb4188) represent genes that participate in the regeneration, production and receptor function of chromophore, which is important for the formation of rhodopsin. Furthermore, the spatial method identified three of the five outliers (ahb8266, est7297, est2423) were associated with the environmental variable insolation, suggesting that these SNPs mark genomic regions that might be involved in local vision-associated adaptation. Three outlier SNPs (est10016, ahb1245 and est5302) mark genes involved in xenobiotic detoxification. These genes have been implicated in tolerance to plant toxins and insecticide resistance. Signal of selection in these genomic regions makes sense because honey bees are exposed to a wide range of natural (plant toxin) or synthetic (agricultural pesticide) xenobiotics during foraging. Furthermore, two of these three SNPs (est10016 and est5302) showed significant correlations with agricultural land cover (a surrogate of agrochemicals use), which suggest the impact of agricultural activities in beekeeping. Finally, two outlier SNPs (est11018 and ahb6903) were related to innate immune response, which participate in the phagocytosis process. Signals of selection in genes related to immune response might be a consequence of the increasing number of emerging honey bee pathogens, due to worldwide movement of honey bee colonies. The spatial analysis corroborated that selection is an underlying force shaping a latitudinal gradient, but it also supported a longitudinal gradient associated with insolation.

Once signals of selection that would be shaping Iberian honey bee variation were identified, analysis of population structure was performed to compare maternal and nuclear patterns. The maternal structure was inferred from polymorphisms in $tRNA^{leu}$-cox2 intergenic region. Mitotypes belonging to two highly divergent A and M lineages were identified in Iberia. The geographical distribution of these mitotypes revealed a sharp southwestern-northeastern clinal pattern. Spatial and non-spatial methods were used to estimate nuclear population structure using the 383 SNPs (complete dataset) and 309 neutrally-tested SNPs (dataset without loci under selection-neutral dataset). Two clusters that remarkably overlapped with the distribution of A and M maternal lineages were detected by those methods. Analysis of geographic clines revealed a clinal pattern for 33 SNPs (17 neutral and 16 loci under selection), as well as for complete and neutral datasets, and mtDNA. Coincidence analysis identified that 18 SNPs (9 neutral and 9 loci under selection), complete and neutral datasets, and mtDNA shareg a common center. The clustering analysis

together with the geographic cline analysis supports a hypothesis of secondary contact between divergent populations. Four aspects of the results support this hypothesis: 1) when comparing the population structure inferred from the 383 (complete dataset) and 309 SNPs, no abrupt changes were observed in the clinal pattern of variation, 2) the presence of several neutral loci forming clines and sharing a common cline center, 3) a trend of elevated values of linkage disequilibrium and diversity patterns towards the center of the cline, and 4) clines formed by mtDNA, complete and neutral datasets, and several individual loci presented coincident centers. Furthermore, the results obtained and based on comparative phylogeography for other Iberian taxa suggest the existence of two putative glacial refuges. The impact of human-mediated processes, as introduction of commercial queens from C lineage and transhumance, is seemingly minimal in Iberian honey bees. This is reflected by the marked clinal pattern at both mitochondrial and nuclear levels and by the presence of only one colony carrying a C-derived mitotype. Intense honey bee transhumance, typical of Iberia, and the high dispersion capacity of honey bees should have altered the clinal pattern, but its persistence could be explained by transhumance occurring after the reproductive season. Local adaptation or some kind of reproductive barrier could also be preventing gene flow and long-term establishment of translocated colonies.

For the first time, a fine-scale study of local adaptation and population structure was carried out for the Iberian honey bee using multiple analytical approaches of population genomics and landscape genetics. The results obtained herein highlight the importance of the Iberian honey bee as a model organism to develop studies of hybrid zones., Additional genomic studies incorporating paleontological and paleoclimatological records, as well as a detailed identification of genomic regions influenced by effects of natural selection or human-mediated processes complemented with studies of functional genomics, are however required to further understand the processes underlying Iberian honey bee variation. Nonetheless, the results reported herein represent an important first step to the understanding of the complex diversity pattern and local adaption of the Iberian honey bee, which can help defining conservation policies of this resource of unquestionable value for the mankind.

Chapter VIII

Supporting material and published papers

Supporting material

Supporting information: Chapter III

Table Sup III - 1. Description of the sample collection

Location (district)	Sample size
Aveiro	31
Beja	45
Braga	26
Bragança	89
Castelo Branco	67
Coimbra	23
Évora	51
Faro	75
Guarda	30
Leiria	67
Lisboa	15
Portalegre	44
Porto	15
Santarém	59
Setúbal	30
Viana do Castelo	32
Vila Real	64
Viseu	23
Azores archipelago	
São Miguel	35
Santa Maria	8
Graciosa	5
Faial	6
Pico	15

Terceira	33
São Jorge	6
Flores	5
Madeira archipelago	
Madeira	51

Supporting information: Chapter IV

Table Sup IV - 1. Description of haplotypes (novel and previously reported) sampled for *A. m. iberiensis* and reference subspecies. Haplotypes previously reported (with or without sequences available in GenBank) are indicated in bold.

Haplotype	Sequence	Fragments	Size
Apis mellifera iberiensis			
Lineage A			
A1	**P_0Q**	**47/108/483**	**638**
A1f	P_0Q	47/108/481	636
A1g	P_0Q	47/108/484	639
A1h	P_0Q	47/108/482	637
A1i	P_0Q	47/109/483	639
A1j	P_0Q	47/108/485	640
A1k	P_0Q	47/108/483	638
A1l	P_0Q	47/108/482	637
A1m	P_0Q	47/108/482	637
A2	**P_0QQ**	**47/108/674**	**829**
A2a	P_0QQ	47/108/672	827
A2b	P_0QQ	47/108/671	826
A2c	P_0QQ	47/108/675	830
A2d	P_0QQ	47/108/675	830
A2e	P_0QQ	47/107/675	829
A2f	P_0QQ	47/108/673	828
A2g	P_0QQ	47/108/672	827
A2h	P_0QQ	47/108/673	828
A2i	P_0QQ	47/108/674	829
A2j	P_0QQ	47/108/678	833
A2k	P_0QQ	47/108/674	829
A2l	P_0QQ	47/108/674	829
A2m	P_0QQ	47/108/673	828
A2n	P_0QQ	47/108/674	829

A2o	P_0QQ	47/108/674	829
A2p	P_0QQ	47/108/673	828
A2q	P_0QQ	47/108/675	830
A2r	P_0QQ	47/108/676	831
A2s	P_0QQ	47/108/672	827
A2t	P_0QQ	47/108/672	827
A2u	P_0QQ	47/108/671	826
A6	**P_0QQ**	**47/108/191/63/420**	**829**
A48	P_0QQ	47/110/674	831
A48a	P_0QQ	47/110/677	834
A48b	P_0QQ	47/110/673	830
A48c	P_0QQ	47/111/674	832
A48d	P_0QQ	47/110/674	831
A48e	P_0QQ	47/110/674	831
A49	P_0QQ	47/108/200/483	838
A50	P_0QQ	47/108/95/577	827
A51	P_0QQ	47/98/673	818
A52	P_0QQ	47/108/196/482	833
A4b	P_0QQ	47/108/192/483	830
A4c	P_0QQ	47/108/191/483	829
A4d	P_0QQ	47/108/191/483	829
A4e	P_0QQ	47/108/192/483	830
A4f	P_0QQ	47/108/191/482	828
A4g	P_0QQ	47/107/192/483	829
A4h	P_0QQ	47/107/192/482	828
A4i	P_0QQ	47/108/191/482	828
A4j	P_0QQ	47/108/191/483	829
A4k	P_0QQ	47/108/191/483	829
A4l	P_0QQ	47/108/191/483	829
A4m	P_0QQ	47/108/191/483	829

A4n	P_0QQ	47/108/190/482	827
A3'	**P_0QQQ**	**47/108/867**	**1022**
A3a'	P_0QQQ	47/108/867	1022
A3b'	P_0QQQ	47/108/867	1022
A3c'	P_0QQQ	47/108/867	1022
A3d'	P_0QQQ	47/108/866	1021
A3e'	P_0QQQ	47/108/866	1021
A3f'	P_0QQQ	47/108/871	1026
A3g'	P_0QQQ	47/108/867	1022
A3h'	P_0QQQ	47/108/867	1022
A3i'	P_0QQQ	47/108/862	1017
A3j'	P_0QQQ	47/108/863	1018
A3k'	P_0QQQ	47/108/864	1019
A3l'	P_0QQQ	47/108/862	1017
A4'	**P_0QQQ**	**47/108/192/192/483**	**1022**
A4a'	P_0QQQ	47/108/192/193/483	1023
A4b'	P_0QQQ	47/108/192/193/483	1023
A53'	P_0QQQ	47/108/192/674	1021
A53a'	P_0QQQ	47/108/191/674	1020
A54'	P_0QQQ	47/108/262/477	894
A8	**P_0Q**	**47/591**	**638**
A8a	P_0Q	47/590	637
A8b	P_0Q	47/594	641
A8c	P_0Q	47/591	638
A8i	P_0Q	47/595	642
A8j	P_0Q	47/596	643
A9a	P_0QQ	47/783	830
A9b	P_0QQ	47/784	831
A9c	P_0QQ	47/785	832
A9d	P_0QQ	47/784	831

A9e	P_0QQ	47/784	831
A9f	P_0QQ	47/784	831
A9g	P_0QQ	47/782	829
A9h	P_0QQ	47/781	828
A10'	**P_0QQQ**	**47/974**	**1021**
A20	**P_1Q**	**47/93/483**	**623**
A20a	P_1Q	47/93/483	623
A11	**P_1QQ**	**47/93/676**	**816**
A11a	P_1QQ	47/93/675	815
A11b	P_1QQ	47/93/674	814
A11c	P_1QQ	47/93/675	815
A14a	P_1QQ	47/93/191/482	813
A14b	P_1QQ	47/93/191/483	814
A30	**P_1QQ**	**47/768**	**815**
A30a	P_1QQ	47/768	815
A30b	P_1QQ	47/764	811
A30c	P_1QQ	47/763	810
A33	**P_1QQ**	**47/93/254/420**	**814**
A55	P_1QQ	47/93/194/482	816
A56	P_1QQ	47/93/668	808
A14'	**P_1QQQ**	**47/93/191/191/482**	**1004**
A14a'	P_1QQQ	47/93/191/190/482	1003
A16'	**P_1QQQ**	**47/93/867**	**1007**
A16b'	P_1QQQ	47/93/866	1006
A16c'	P_1QQQ	47/93/866	1006
A16d'	P_1QQQ	47/93/866	1006
A16e'	P_1QQQ	47/93/867	1007
A16f'	P_1QQQ	47/93/868	1008
A16g'	P_1QQQ	47/93/870	1010
A16h'	P_1QQQ	47/92/866	1005

A16i’	P_1QQQ	47/93/865	1005
A16j’	P_1QQQ	47/93/865	1005
A16k’	P_1QQQ	47/93/869	1009
A16l’	P_1QQQ	47/93/867	1007
A39’	**P_1QQQ**	**47/65/28/867**	**1007**
A41a’	P_1QQQ	47/269/674	990
A42a’	**P_1QQQ**	**47/93/447/420**	**1007**
A43b’	P_1QQQ	47/93/192/674	1006
A57’	P_1QQQ	47/93/852	992
A58’	P_1QQQ	47/944	991
A59’	P_1QQQ	47/957	1004
A60’	P_1QQQ	47/93/451/420	1011
A61’	P_1QQQ	47/30/63/866	1006
A62’	P_1QQQ	140/867	1007
	Lineage M		
M79	P_0Q	157/65/422	644
M79a	P_0Q	156/65/422	643
M3a	**PQ**	**47/95/65/422**	**629**
M14	**PQ**	**114/28/65/422**	**629**
M4	**PQQ**	**142/65/131/65/422**	**825**
M4f	PQQ	142/65/131/65/422	825
M4g	PQQ	142/65/131/65/422	825
M4h	PQQ	142/65/130/65/422	824
M4i	PQQ	141/65/131/65/422	824
M4j	PQQ	142/65/131/65/422	825
M4k	PQQ	142/65/130/65/421	824
M4l	PQQ	142/65/131/65/422	825
M4m	PQQ	142/65/132/65/422	826
M4n	PQQ	142/65/131/65/422	825
M4o	PQQ	142/65/131/65/422	825

M4p	PQQ	143/65/131/65/422	826
M17i	PQQ	142/66/131/65/422	826
M19b	PQQ	142/64/131/65/422	824
M41a	PQQ	142/65/131/69/422	829
M42a	PQQ	142/65/131/67/422	827
M70	PQQ	145/65/131/65/422	828
M70a	PQQ	145/65/131/65/422	828
M70b	PQQ	145/65/131/65/422	828
M72	PQQ	142/67/131/67/422	829
M75	PQQ	146/65/131/65/422	829
M8	**PQQ**	**114/28/65/131/65/422**	**825**
M8a	PQQ	114/28/65/131/65/422	825
M8b	PQQ	114/28/65/131/65/422	825
M11a	PQQ	114/28/66/131/65/422	826
M7	**PQQ**	**47/95/65/131/65/422**	**825**
M7g	PQQ	47/94/65/131/65/422	824
M7h	PQQ	47/95/65/131/65/421	825
M7i	PQQ	47/95/65/131/65/422	825
M7j	PQQ	47/95/65/130/65/422	824
M7k	PQQ	47/95/65/131/65/422	825
M7l	PQQ	47/95/65/131/65/422	825
M7m	PQQ	47/95/65/131/65/422	825
M7n	PQQ	47/94/65/131/65/422	824
M7o	PQQ	47/94/65/130/65/422	823
M68	PQQ	47/95/64/131/64/422	823
M69	PQQ	47/96/64/131/65/422	825
M69a	PQQ	47/96/64/131/65/422	825
M69b	PQQ	47/95/65/131/64/422	824
M69c	PQQ	47/95/65/131/64/422	824
M73	PQQ	43/94/65/131/65/422	820

M74	PQQ	47/95/65/131/61/422	821
M76	PQQ	47/94/65/94/37/65/422	824
M78	PQQ	47/94/69/131/64/422	827
M4'	**PQQQ**	**142/65/131/65/131/65/422**	**1021**
M4a'	PQQQ	142/65/131/65/131/65/422	1021
M4b'	PQQQ	142/65/131/65/131/65/422	1021
M4e'	PQQQ	142/65/131/65/131/65/422	1021
M4f'	PQQQ	142/65/131/65/131/65/422	1021
M40'	PQQQ	142/68/131/65/131/65/422	1024
M8'	**PQQQ**	**114/28/65/131/65/131/65/422**	**1021**
M8a'	PQQQ	114/28/65/131/65/131/65/422	1021
M8b'	PQQQ	115/28/65/130/65/130/65/422	1020
M7'	**PQQQ**	**47/95/65/131/65/131/65/422**	**1021**
M77'	PQQQ	99/43/65/131/65/131/65/422	1021
		Lineage C	
C2j	**Q**	**47/40/64/420**	**571**
		Apis mellifera intermissa	
		Lineage A	
A1n	P_0Q	47/108/482	637
A1o	P_0Q	47/108/482	637
A1p	P_0Q	47/109/482	638
A8d	P_0Q	47/590	637
A8e	P_0Q	47/592	639
A8f	P_0Q	47/591	638
A8g	P_0Q	47/591	638
A8h	P_0Q	47/590	637
A9i	P_0QQ	47/782	829
A9j	P_0QQ	47/782	829
A9k	P_0QQ	47/783	830
A9l	P_0QQ	47/783	830

A9m	P_0QQ	47/781	828
A9n	P_0QQ	47/781	828
A9o	P_0QQ	47/782	829
A9p	P_0QQ	47/782	829
A9q	P_0QQ	47/780	827
A63	P_0QQ	47/765	812
Apis mellifera mellifera			
Lineage M			
M4	**PQQ**	**142/65/131/65/422**	**825**
M17c	**PQQ**	**142/66/131/65/422**	**826**
M17d	**PQQ**	**142/66/131/65/422**	**826**
M17e	**PQQ**	**142/66/131/65/422**	**826**
M17f	**PQQ**	**142/66/131/65/422**	**826**
M17h	**PQQ**	**142/66/131/65/421**	**825**
M71	**PQQ**	**142/66/131/66/422**	**827**
M17a'	**PQQQ**	**142/66/131/65/131/65/422**	**1022**
Apis mellifera ligústica			
Lineage M			
M7a	**PQQ**	**47/95/65/131/65/422**	**825**
Lineage C			
C1	**Q**	**47/41/64/420**	**572**
Apis mellifera carnica			
Lineage C			
C1	**Q**	**47/41/64/420**	**572**
C1b	**Q**	**47/41/64/420**	**572**
C2	**Q**	**47/40/64/420**	**571**
C2c	**Q**	**47/40/64/420**	**571**
C3	**Q**	**47/40/63/420**	**570**
C3a	**Q**	**47/40/63/420**	**570**

Table Sup IV - 2. List of previously reported haplotypes with description updated. (') Haplotypes described without sequences available in GenBank, (ᵃ) haplotypes renamed or with *Dra*I band pattern and fragment size amended.

Haplotype	Sequence	Fragments	Size	Accession number	Reference	Novel haplotype	Fragments	Size
Lineage A								
A1b[&]	P_0Q	47/108/483	638	FJ477985.1	Franck *et al*. 2001	A1	47/108/483	638
A1[&]	P_0Q	47/108/483	638	EF033649.1	Collet *et al*. 2006	A1b	47/108/483	638
A1[&]	P_0Q	47/108/483	638	KJ661741.1	Branchiccela *et al*. 2014	A1e	47/108/483	638
A2[*]	P_0QQ	47/108/676	831	-	Franck *et al*. 2001	A2	47/108/674	829
A3[*&]	P_0QQQ	47/108/869	1024	-	Franck *et al*. 2001	A3'	47/108/867	1022
A4	P_0QQ	47/108/192/483	830	EF033650.1	Collet *et al*. 2006	A4	47/108/192/483	830
A4[&]	P_0QQ	47/107/191/483	828	FJ477987.1	Franck *et al*. 2001	A4a	47/107/191/483	828
A6[*]	P_0QQ	47/108/191/64/420	830	-	Franck *et al*. 2001	A6	47/108/191/63/420	829
A4'[*]	P_0QQQ	47/108/193/193/483	1024	-	Franck *et al*. 2001	A4'	47/108/192/192/483	1022
A46'	P_0QQQ	47/108/267/600	1022	JQ746701.1	Pinto *et al*. 2012	A46'	47/108/267/600	1022
A8	P_0Q	47/591	638	FJ477981.1	Franck *et al*. 2001	A8	47/591	638
A9[&]	P_0QQ	47/784	831	FJ477982.1	Franck *et al*. 2001	A9	47/783	830
A10[*&]	P_0QQQ	47/977	1024	-	Franck *et al*. 2001	A10'	47/974	1021
A20[*]	P_1Q	47/93/483	623	-	Franck *et al*. 2001	A20	47/93/483	623
A11[*]	P_1QQ	47/93/676	816	-	Franck *et al*. 2001	A11	47/93/676	816
A14[&]	P_1QQ	47/93/193/483	816	FJ477991.1	Franck *et al*. 2001	A14	47/93/190/484	815
A30	P_1QQ	47/768	815	EF033654.1	Collet *et al*. 2006	A30	47/768	815
A33	P_1QQ	47/93/254/420	814	JQ746686.1	Pinto *et al*. 2012	A33	47/93/254/420	814
A36	P_1QQ	47/65/28/676	816	JQ746689.1	Pinto *et al*. 2012	A36	47/65/28/676	816

A37	P_1QQ	47/65/28/192/483	815	JQ746690.1	Pinto *et al*. 2012	A37	47/65/28/192/483	815
A15*&	P_1QQQ	47/93/193/193/483	1009	-	Franck *et al*. 2001	A14’	47/93/191/191/482	1004
A16*&	P_1QQQ	47/93/869	1009	-	Franck *et al*. 2001	A16’	47/93/867	1007
A39’	P_1QQQ	47/65/28/867	1007	JQ746692.1	Pinto *et al*. 2012	A39’	47/65/28/867	1007
A41’	P_1QQQ	47/270/675	992	JQ746694.1	Pinto *et al*. 2012	A41’	47/270/675	992
A42’	P_1QQQ	47/93/447/420	1007	JQ746695.1	Pinto *et al*. 2012	A42’	47/93/447/420	1007
A42a’	P_1QQQ	47/93/447/420	1007	JQ746695.1	Pinto *et al*. 2012	A42a’	47/93/447/420	1007
A43’	P_1QQQ	47/93/192/673	1005	JQ746697.1	Pinto *et al*. 2012	A43’	47/93/192/673	1005
A47&	P_1QQQQ	47/93/191/191/191/482	1195	JX898863.1	Muñoz *et al*. 2013	A14’’	47/93/191/191/191/482	1195
Z3	P_0Q	47/108/66/420	641	HM236207.1	Alburaki *et al*. 2011	Z3	47/108/66/420	641
Z7	P_0Q	47/108/67/420	642	HM236204.1	Alburaki *et al*. 2011	Z7	47/108/67/420	642
Z21	P_0Q	47/108/67/86/334	642	HM236218.1	Alburaki *et al*. 2011	Z21	47/108/67/86/334	642
Z1	P_0QQ	47/108/67/114/67/420	823	HM236204.1	Alburaki *et al*. 2011	Z1	47/108/67/114/67/420	823
Z2	P_0QQ	47/108/67/129/67/420	838	HM236206.1	Alburaki *et al*. 2011	Z2	47/108/67/129/67/420	838
Lineage M								
M12&	P_0Q	157/65/422	644	HQ337445.1	Rortais *et al*. 2011	M12	157/65/87/335	644
M13&	P_0Q	157/67/422	646	HQ337446.1	Rortais *et al*. 2011	M13	157/67/423	647
M63&	P_0Q	152/61/130/61/422	826	HM236202.1	Alburaki *et al*. 2011	M63	156/61/421	638
M10&	P_0QQ	157/67/130/65/422	841	HQ337443.1	Rortais *et al*. 2011	M10	157/67/130/65/423	842
M10a’	P_0QQQ	157/67/131/65/131/65/422	1038	HQ337444.1	Rortais *et al*. 2011	M10a’	157/67/131/65/131/65/422	1038
M16’	P_0QQQ	157/65/131/65/116/65/422	1021	HQ337449.1	Rortais *et al*. 2011	M16’	157/65/131/65/116/65/422	1021
M3	PQ	47/94/65/422	628	HQ337435.1	Rortais *et al*. 2011	M3	47/94/65/422	628

M6	PQ	142/65/422	629	FJ478008.1	Franck *et al*.2001	M6	142/65/422	629
M30	PQ	142/66/422	630	HQ260340.1	Rortais *et al*. 2011	M30	142/66/422	630
M62[&]	PQ	142/64/422	628	HQ260341.1	Rortais *et al*. 2011	M62	142/64/423	629
M14	PQ	114/28/65/422	629	HQ337447.1	Rortais *et al*. 2011	M14	114/28/65/422	629
M4a[&]	PQQ	142/65/131/65/422	825	KF274625.1	Pinto *et al*. 2014	M4	142/65/131/65/422	825
M4[&]	PQQ	142/65/131/65/422	825	FJ478006.1	Franck *et al*. 2001	M4a	142/65/131/65/422	825
M4g[&]	PQQ	142/65/131/65/424	827	KF274630.1	Pinto *et al*. 2014	M4b	142/65/131/65/424	827
M4k[&]	PQQ	142/65/131/65/422	825	KF274634.1	Pinto *et al*. 2014	M4c	142/65/131/65/422	825
M4m[&]	PQQ	142/65/131/65/422	825	KF274636.1	Pinto *et al*. 2014	M4d	142/65/131/65/422	825
M4n[&]	PQQ	142/65/130/65/422	824	KF274637.1	Pinto *et al*. 2014	M4e	142/65/130/65/422	824
M17	PQQ	142/66/131/65/422	826	HQ337450.1	Rortais *et al*. 2011	M17	142/66/131/65/422	826
M4[&]	PQQ	141/66/131/65/422	825	EF033656.1	Collet *et al*. 2006	M17b	141/66/131/65/422	825
M4b[&]	PQQ	142/66/131/65/422	826	KF274626.1	Pinto *et al*. 2014	M17c	142/66/131/65/422	826
M4d[&]	PQQ	142/66/131/65/422	826	KF274627.1	Pinto *et al*. 2014	M17d	142/66/131/65/422	826
M4e[&]	PQQ	142/66/131/65/422	826	KF274628.1	Pinto *et al*. 2014	M17e	142/66/131/65/422	826
M4f[&]	PQQ	142/66/131/65/422	826	KF274629.1	Pinto *et al*. 2014	M17f	142/66/131/65/422	826
M4j[&]	PQQ	142/66/131/65/421	825	KF274633.1	Pinto *et al*. 2014	M17h	142/66/131/65/421	825
M4i[&]	PQQ	142/66/131/66/422	827	KF274632.1	Pinto *et al*. 2014	M71	142/66/131/66/422	827
M42	PQQ	142/67/131/65/422	827	HQ260353.1	Rortais *et al*. 2011	M42	142/67/131/65/422	827
M64	PQQ	142/65/130/30/65/422	854	KF274640.1	Pinto *et al*. 2014	M64	142/65/130/30/65/422	854
M8	PQQ	114/28/65/131/65/422	825	FJ478007.1	Franck *et al*. 2001	M8	114/28/65/131/65/422	825
M11[&]	PQQ	114/28/66/131/65/422	827	HQ260343.1	Rortais *et al*. 2011	M11	114/28/66/131/65/423	827

M15	PQQ	114/28/65/125/6/65/422	825	HQ337448.1	Rortais *et al*. 2011	M15	114/28/65/125/6/65/422	825
M34	PQQ	114/28/64/131/65/422	824	HQ337456.1	Rortais *et al*. 2011	M34	114/28/64/131/65/423	825
M7	PQQ	47/95/65/131/65/422	825	FJ478005.1	Franck *et al*. 2001	M7	47/95/65/131/65/422	825
M7a	PQQ	47/95/65/131/65/422	825	KF274639.1	Pinto *et al*. 2014	M7a	47/95/65/131/65/422	825
M4'*	PQQQ	142/65/131/65/131/65/422	1021	-	Franck *et al*. 2001	M4'	142/65/131/65/131/65/422	1021
M4a'&	PQQQ	142/66/131/65/131/65/422	1022	KF274638.1	Pinto *et al*. 2014	M17a'	142/66/131/65/131/65/422	1022
M39'	PQQQ	142/61/131/61/131/65/422	1013	HQ260369.1	Rortais *et al*. 2011	M39'	142/61/131/61/131/65/422	1013
M55'&	PQQQ	142/64/128/67/131/67/422	1021	HQ260371.1	Rortais *et al*. 2011	M55'	142/64/128/67/131/67/423	1022
M57'	PQQQ	142/64/129/65/128/65/422	1015	HQ260372.1	Rortais *et al*. 2011	M57'	142/64/129/65/128/65/422	1015
M8'	PQQQ	114/28/65/131/65/131/65/422	1021	HQ337441.1	Rortais *et al*. 2011	M8'	114/28/65/131/65/131/65/422	1021
M7'*	PQQQ	47/95/65/131/65/131/65/422	1022	-	Rortais *et al*. 2011	M7'	47/95/65/131/65/131/65/422	1021
M58'	PQQQ	47/95/65/131/65/136/422	961	HQ260373.1	Rortais *et al*. 2011	M58'	47/95/65/131/65/136/422	961
Lineage C								
C1	Q	47/41/64/420	572	FJ478010.1	Franck *et al*. 2001	C1	47/41/64/420	572
C1b	Q	47/41/64/420	572	FJ357799.1	Ödzil *et al*. 2009	C1b	47/41/64/420	572
C2*	Q	47/40/64/420	571	-	Franck *et al*. 2001	C2	47/40/64/420	571
C2d&	Q	47/40/64/420	571	FJ824584.1	Muñoz *et al*. 2009	C2	47/40/64/420	571
C2a	Q	47/40/64/420	571	FJ357805.1	Franck *et al*. 2001	C2a	47/40/64/420	571
C2c	Q	47/40/64/420	571	FJ824583.1	Muñoz *et al*. 2009	C2c	47/40/64/420	571
C2j	Q	47/40/64/420	571	HQ199228.1	Muñoz and De la Rúa 2012	C2j	47/40/64/420	571
C2&	Q	47/40/64/420	571	KF955994.1	Techer *et al*. 2015	C2j	47/40/64/420	571
C11&	Q	47/40/64/420	571	FJ037776.1	Solórzano *et al*. 2009	C2j	47/40/64/420	571

C3*	Q	47/40/63/420	570	-	Perrier *et al*. 2003	C3	47/40/63/420	570
C2e&	Q	47/40/63/420	570	FJ824586.1	Muñoz *et al*. 2009	C3	47/40/63/420	570
C2k&	Q	47/40/63/420	570	GQ433624.1	Razpet *et al*. unpublished	C3a	47/40/63/420	570
C31&	Q	47/40/63/420	570	HQ287900.1	Magnus *et al*. 2011	C3b	47/40/63/420	570
C8	Q	47/40/64/410	561	HM236203.1	Alburaki *et al*. 2011	C8	47/40/64/410	561
Isolate C34&	Q	47/40/60/420	567	JF934710.1	Magnus *et al*. 2011	C34	47/40/60/420	567
C21	Q	47/40/64/415	566	FJ037784.1	Solórzano *et al*. 2009	C21	47/40/64/415	566

Table Sup IV - 3. Haplotype frequencies and sample sizes (on the second line) for each Iberian sampling site (AT - Atlantic transect, CT - central transect, MT - Mediterranean transect) and reference subspecies (CAR - *A. m. carnica*, LIG - *A. m. ligustica*, MEL - A. *m. mellifera*, INT - *A. m. intermissa*).

	AT1	**AT2**	**AT3**	**AT4**	**AT5**	**AT6**	**AT7**	**AT8**	**CT1**	**CT2**	**CT3**	**CT4**	**CT5**	**CT6**	**CT7**	**CT8**	**CT9**	**MT1**	**MT2**	**MT3**	**MT4**	**MT5**	**MT6**	**INT**	**MEL**	**LIG**	**CAR**
Haplotypes	30	30	30	30	33	39	30	30	30	30	30	30	30	33	30	30	24	36	30	30	33	33	30	31	34	17	19
A49	0.133	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M72	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A11	0.033	0.100	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M7g	0.067	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M69c	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M42a	0.033	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M7h	0.067	0.067	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M74	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M68	0.067	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M7	0.133	0.067	0.000	0.000	0.000	0.000	0.000	0.000	0.067	0.033	0.133	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000
M41a	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M7o	0.067	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M7j	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M79a	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.030	0.000	0.000	0.000	0.000	0.000	0.000
A2a	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.133	0.091	0.033	0.033	0.000	0.083	0.000	0.000	0.333	0.333	0.333	0.000	0.000	0.000	0.000
M7i	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A8	0.033	0.000	0.000	0.067	0.030	0.051	0.133	0.000	0.067	0.000	0.000	0.400	0.267	0.061	0.033	0.000	0.000	0.000	0.100	0.000	0.000	0.061	0.000	0.000	0.000	0.000	0.000
M73	0.100	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A16'	0.000	0.200	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A4c	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A4d	0.000	0.067	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A1f	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A2c	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.061	0.000	0.267	0.083	0.028	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M7k	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M78	0.000	0.100	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A3h'	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A20a	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

A1	0.000	0.033	0.000	0.000	0.091	0.179	0.167	0.733	0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A3c'	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A58'	0.000	0.100	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A16i'	0.000	0.000	0.067	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A16c'	0.000	0.000	0.100	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A14a	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M4	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.300	0.433	0.033	0.133	0.000	0.000	0.100	0.000	0.000	0.222	0.367	0.300	0.242	0.061	0.000	0.000	0.029	0.000	0.000
A1g	0.000	0.000	0.067	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A4a'	0.000	0.000	0.067	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
C2j	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A16b'	0.000	0.000	0.167	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A11b	0.000	0.000	0.067	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A16d'	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A4e	0.000	0.000	0.067	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A4b	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A16f'	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A30	0.000	0.000	0.033	0.067	0.030	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A9a	0.000	0.000	0.067	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.042	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A3d'	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A16e'	0.000	0.000	0.033	0.000	0.000	0.026	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A8j	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A30b	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A43b'	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A16j'	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A62'	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A1h	0.000	0.000	0.000	0.200	0.091	0.000	0.033	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.042	0.000	0.000	0.100	0.030	0.152	0.167	0.000	0.000	0.000	0.000
A30c	0.000	0.000	0.000	0.067	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A53'	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A55	0.000	0.000	0.000	0.067	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A14b	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A2d	0.000	0.000	0.000	0.033	0.000	0.077	0.100	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

A16l'	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A39'	0.000	0.000	0.000	0.067	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A30a	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A8a	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A57'	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A59'	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A16g'	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A20	0.000	0.000	0.000	0.000	0.242	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A2e	0.000	0.000	0.000	0.000	0.061	0.000	0.133	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A3a'	0.000	0.000	0.000	0.000	0.121	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A4f	0.000	0.000	0.000	0.000	0.061	0.026	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A33	0.000	0.000	0.000	0.000	0.030	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A3e'	0.000	0.000	0.000	0.000	0.030	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A9b	0.000	0.000	0.000	0.000	0.030	0.000	0.000	0.067	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A10'	0.000	0.000	0.000	0.000	0.061	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A4g	0.000	0.000	0.000	0.000	0.091	0.026	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A56	0.000	0.000	0.000	0.000	0.030	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A4h	0.000	0.000	0.000	0.000	0.000	0.103	0.100	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A8b	0.000	0.000	0.000	0.000	0.000	0.077	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A2h	0.000	0.000	0.000	0.000	0.000	0.051	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A14'	0.000	0.000	0.000	0.000	0.000	0.051	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A42a'	0.000	0.000	0.000	0.000	0.000	0.077	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A11c	0.000	0.000	0.000	0.000	0.000	0.026	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A9c	0.000	0.000	0.000	0.000	0.000	0.077	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A16h'	0.000	0.000	0.000	0.000	0.000	0.026	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A60'	0.000	0.000	0.000	0.000	0.000	0.051	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A4b'	0.000	0.000	0.000	0.000	0.000	0.026	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A3f'	0.000	0.000	0.000	0.000	0.000	0.026	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A1i	0.000	0.000	0.000	0.000	0.000	0.026	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A2i	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A1j	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

A16k'	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A9d	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A2j	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A8i	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A3'	0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.000	0.000	0.000	0.000	0.167	0.067	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A1k	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.067	0.000	0.000	0.067	0.033	0.033	0.273	0.133	0.267	0.417	0.000	0.000	0.067	0.000	0.030	0.233	0.000	0.000	0.000	0.000
M7n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.200	0.000	0.000	0.000	0.061	0.000	0.000	0.000	0.000	0.133	0.000	0.061	0.000	0.000	0.000	0.000	0.000	0.000
M4f	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.067	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M4g	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.167	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M69	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.067	0.067	0.000	0.000	0.133	0.000	0.000	0.000	0.000	0.028	0.033	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M4p	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.067	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M4h	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.067	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A2g	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.033	0.061	0.000	0.000	0.000	0.000	0.000	0.000	0.030	0.030	0.067	0.000	0.000	0.000	0.000
A50	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.067	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M4a'	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M4'	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.056	0.067	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M40'	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.061	0.000	0.000	0.000	0.000	0.000
M8a'	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A2s	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.067	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A9e	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M75	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.133	0.000	0.033	0.000	0.000	0.000	0.000	0.222	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M3a	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M7'	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M79	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.267	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M11a	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.067	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M7l	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A41a'	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A2f	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.067	0.000	0.000	0.067	0.000	0.042	0.000	0.000	0.000	0.030	0.000	0.000	0.000	0.000	0.000	0.000
A11a	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.067	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

A4i	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A9f	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A48b	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.067	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A4n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.067	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.030	0.000	0.000	0.000	0.000	0.000
A2k	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.033	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A8c	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.067	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A48a	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.030	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A3l	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.091	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A48c	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.091	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A4j	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.030	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A3k	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.091	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A2l	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.030	0.000	0.067	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A2m	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.030	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A9g	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.033	0.042	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A4_l	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A48	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.067	0.000	0.083	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A14a'	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A4'	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A53a'	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.067	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A1_l	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A4k	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A2n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A4m	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A48d	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.042	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A51	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.042	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A48e	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.083	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A2r	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.083	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M4i	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.028	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M4j	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.083	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M4k	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.028	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

M8a	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.028	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M76	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.056	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M8'	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.083	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M14	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.028	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M4b'	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.028	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A3j'	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M4l	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M19b	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M69b	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.067	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M8b	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M4m	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M4n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M77'	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M4e'	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.067	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M4o	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.030	0.000	0.000	0.000	0.000	0.000	0.000
M70b	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M4f'	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M70	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M8b'	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M70a	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M69a	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A61'	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.030	0.000	0.000	0.000	0.000	0.000	0.000
A54'	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.091	0.000	0.000	0.000	0.000	0.000	0.000
A9h	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.030	0.000	0.000	0.000	0.000	0.000	0.000
M17i	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.030	0.000	0.000	0.000	0.000	0.000	0.000
A2o	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.030	0.000	0.000	0.000	0.000	0.000	0.000
A3i'	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.030	0.000	0.000	0.000	0.000	0.000	0.000
A2p	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.030	0.000	0.000	0.000	0.000	0.000	0.000
M7m	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.030	0.000	0.000	0.000	0.000	0.000	0.000
A2t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.030	0.000	0.000	0.000	0.000	0.000	0.000
A52	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.030	0.000	0.000	0.000	0.000	0.000	0.000

A2u	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.030	0.000	0.000	0.000	0.000	0.000
A2b	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.030	0.000	0.000	0.000	0.000	0.000
A2q	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.030	0.000	0.000	0.000	0.000	0.000
A3b'	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.000	0.000	0.000	0.000
A3g'	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000
A1m	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000
A9i	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.097	0.000	0.000	0.000	0.000
A8d	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.065	0.000	0.000	0.000	0.000
A9j	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.065	0.000	0.000	0.000	0.000
A9k	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.097	0.000	0.000	0.000	0.000
A8e	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.032	0.000	0.000	0.000	0.000
A9p	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.032	0.000	0.000	0.000	0.000
A9q	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.032	0.000	0.000	0.000	0.000
A8f	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.065	0.000	0.000	0.000	0.000
A8g	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.032	0.000	0.000	0.000	0.000
A9m	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.032	0.000	0.000	0.000	0.000
A9l	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.032	0.000	0.000	0.000	0.000
A9o	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.032	0.000	0.000	0.000	0.000
A8h	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.129	0.000	0.000	0.000	0.000
A1p	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.065	0.000	0.000	0.000	0.000
A1n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.032	0.000	0.000	0.000	0.000
A63	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.065	0.000	0.000	0.000	0.000
A1o	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.065	0.000	0.000	0.000	0.000
A9n	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.032	0.000	0.000	0.000	0.000
M17b	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.118	0.000	0.000	0.000
M17c	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.412	0.000	0.000	0.000
M17e	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.029	0.000	0.000	0.000
M17d	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.059	0.000	0.000	0.000
M17a'	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.265	0.000	0.000	0.000
M71	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.059	0.000	0.000	0.000
M17g	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.029	0.000	0.000	0.000

C1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.706	0.053
M7a	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.294	0.000
C3a	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.053
C3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.316
C2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.421
C2c	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.105
C1b	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.053

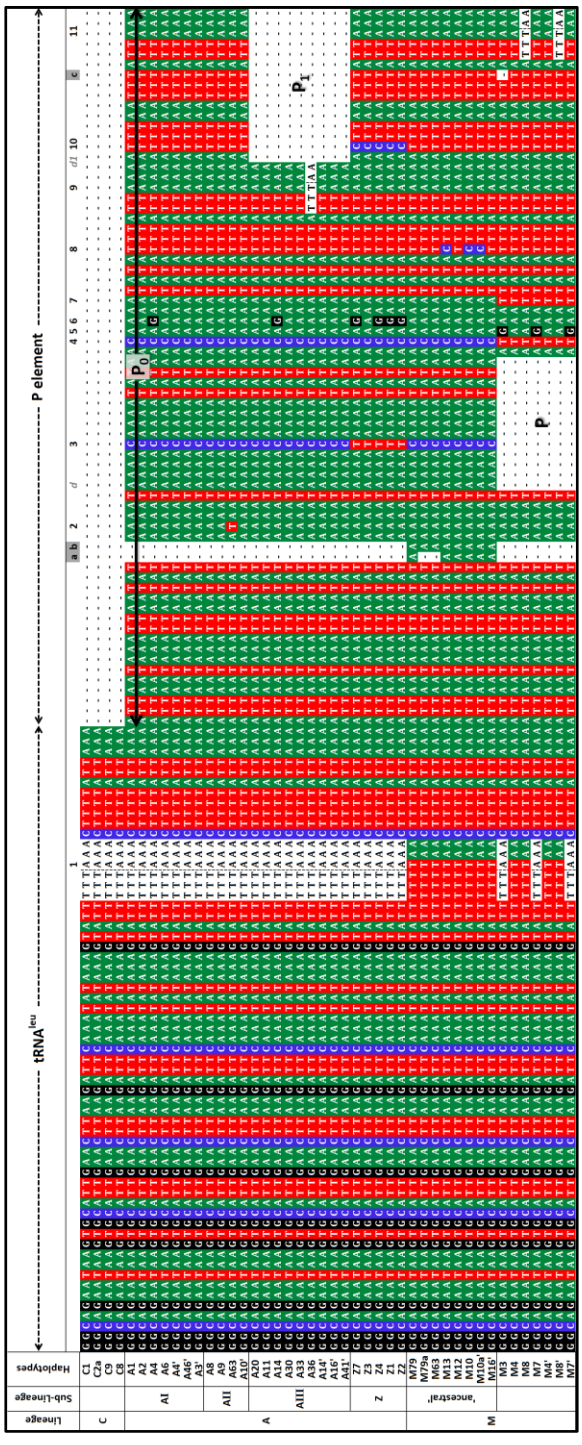


Figure Sup IV - 1. Sequence alignment of the tRNA^{Leu}-cox2 intergenic region showing variation among haplotypes of lineage C, A, and M ancestry. Substitution sites are numbered from 1 to 47. Indels (indicated with a dash) are denoted by letters from "a" to "d" with a grey background. *Dra*I recognition sites (TTTAAA) are indicated by nucleotides in bold and white background. The absence of the P element characterizes lineage C. Sub-lineages A, A_I and Z carry the P_O element but they are differentiated by their *Dra*I restriction sites. Sub-lineage A_{III} is identified by the P_1 element (indicated by the "*d*" deletion marked in grey). Haplotypes of lineage M are mostly differentiated by the P element (indicated by the "*d*" deletion marked in grey). However, 'ancestral' M haplotypes present the P_O element and the same restriction sites as other haplotypes of lineage M.

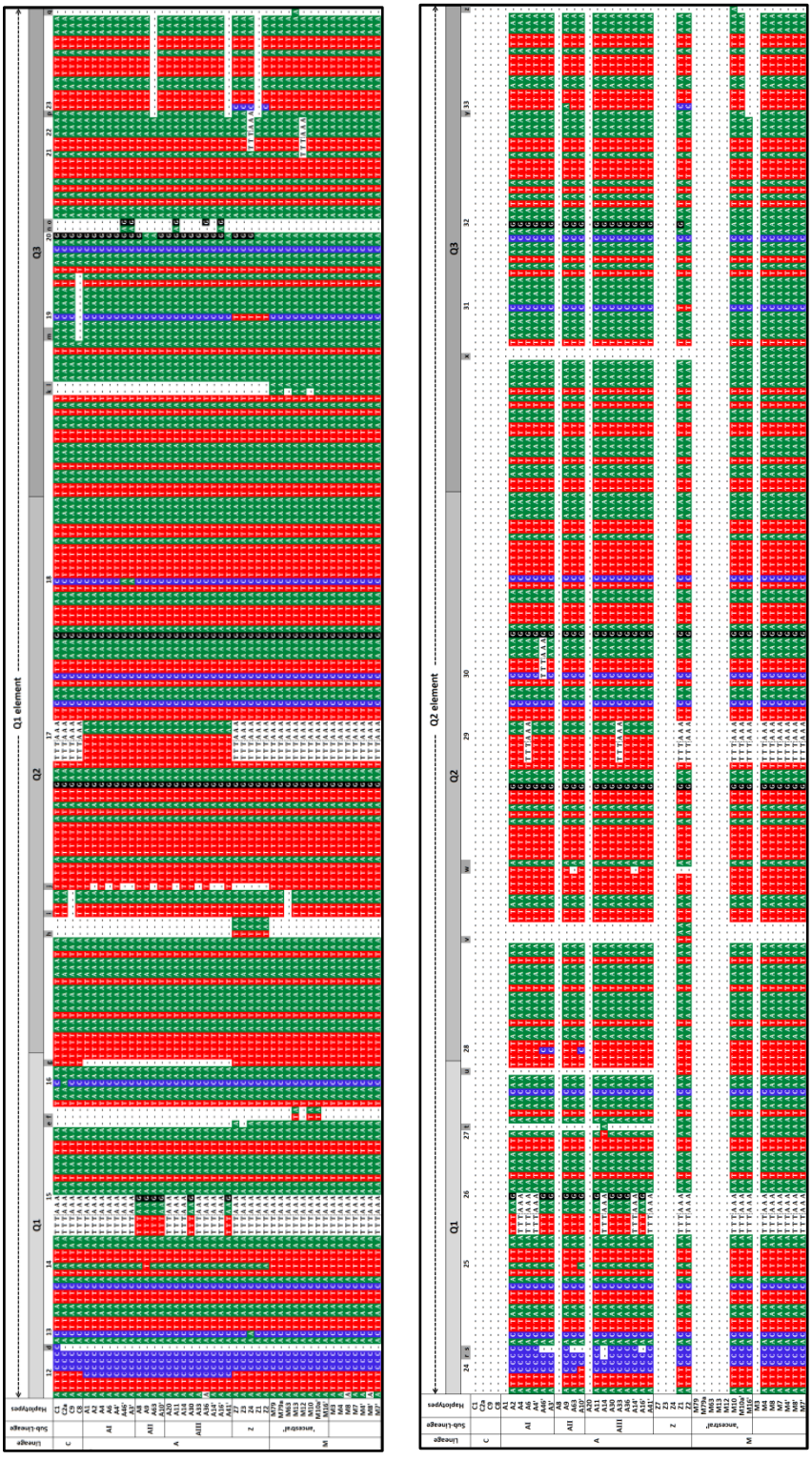


Figure Sup IV – 1. (Cont.) Structure of the Q1 and Q2 elements showing the three distinct portions (Q_1, Q_2, Q_3).

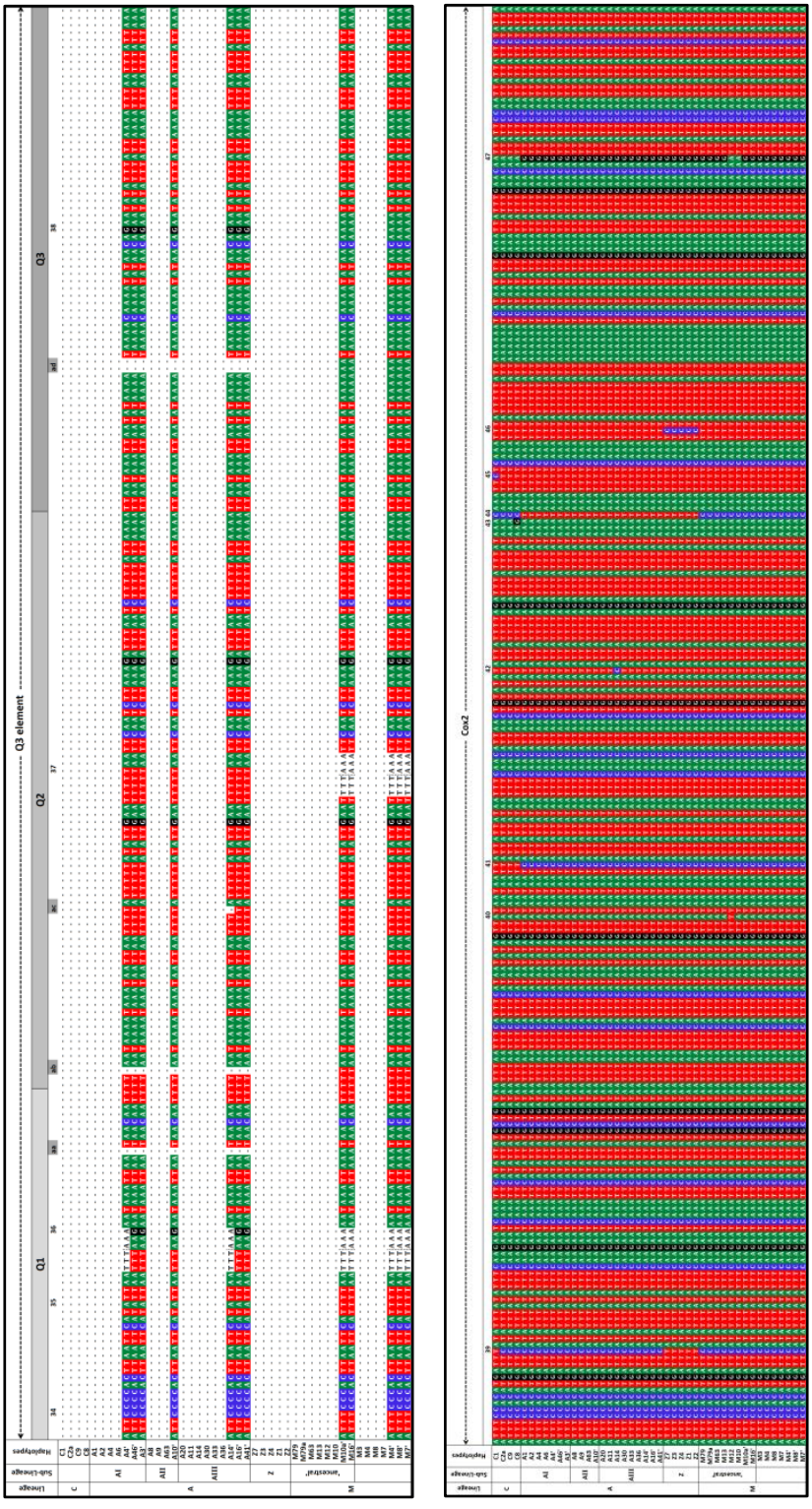


Figure Sup IV – 1. (Cont.) Structure of the Q3 element showing the three distinct portions (Q_1, Q_2, Q_3) and the 5´end of cox2.

Supporting information: Chapter V

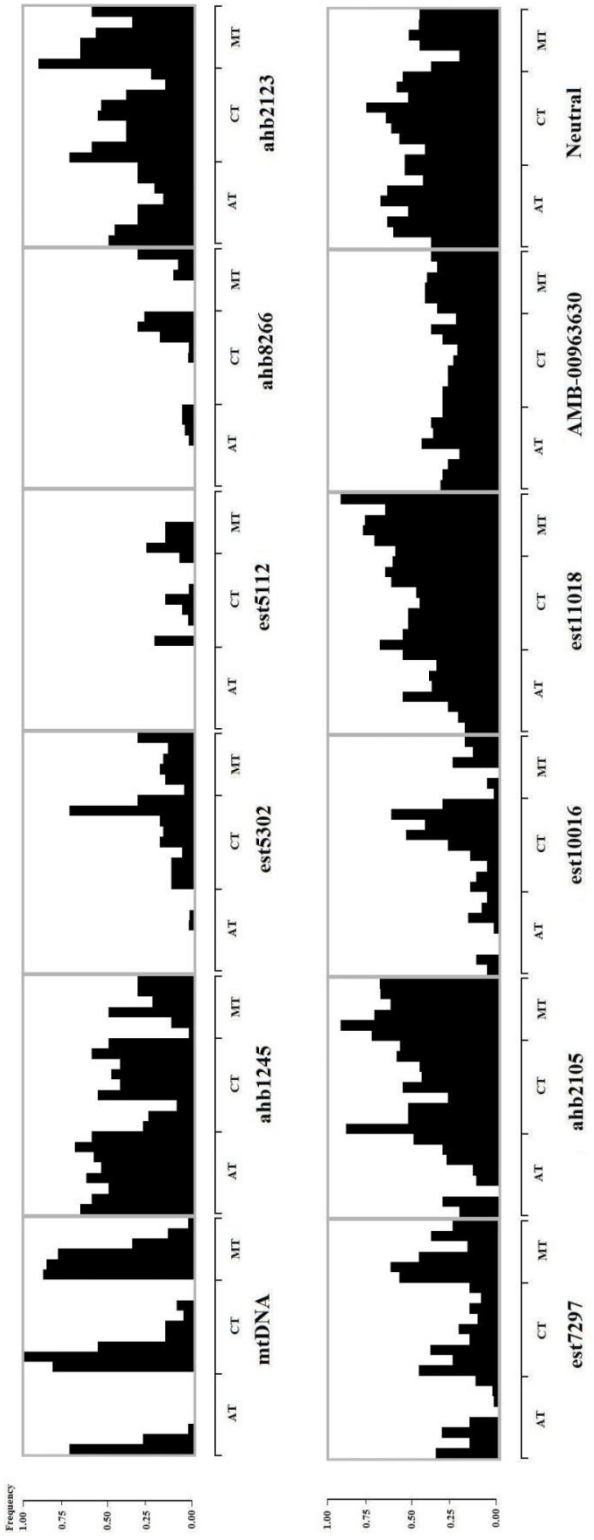


Figure Sup V - 1. Individual bar charts indicate allele frequencies of mtDNA (M-lineage haplotypes), the strongest directional outliers (from ahb1245 to est11018), the strongest balancing outlier (AMB-00963630), and a representative neutral locus, in each of the 23 sampling sites. For each plot, sampling sites from left to right are AT1 to AT8, CT1 to CT9, and MT1 to MT6.

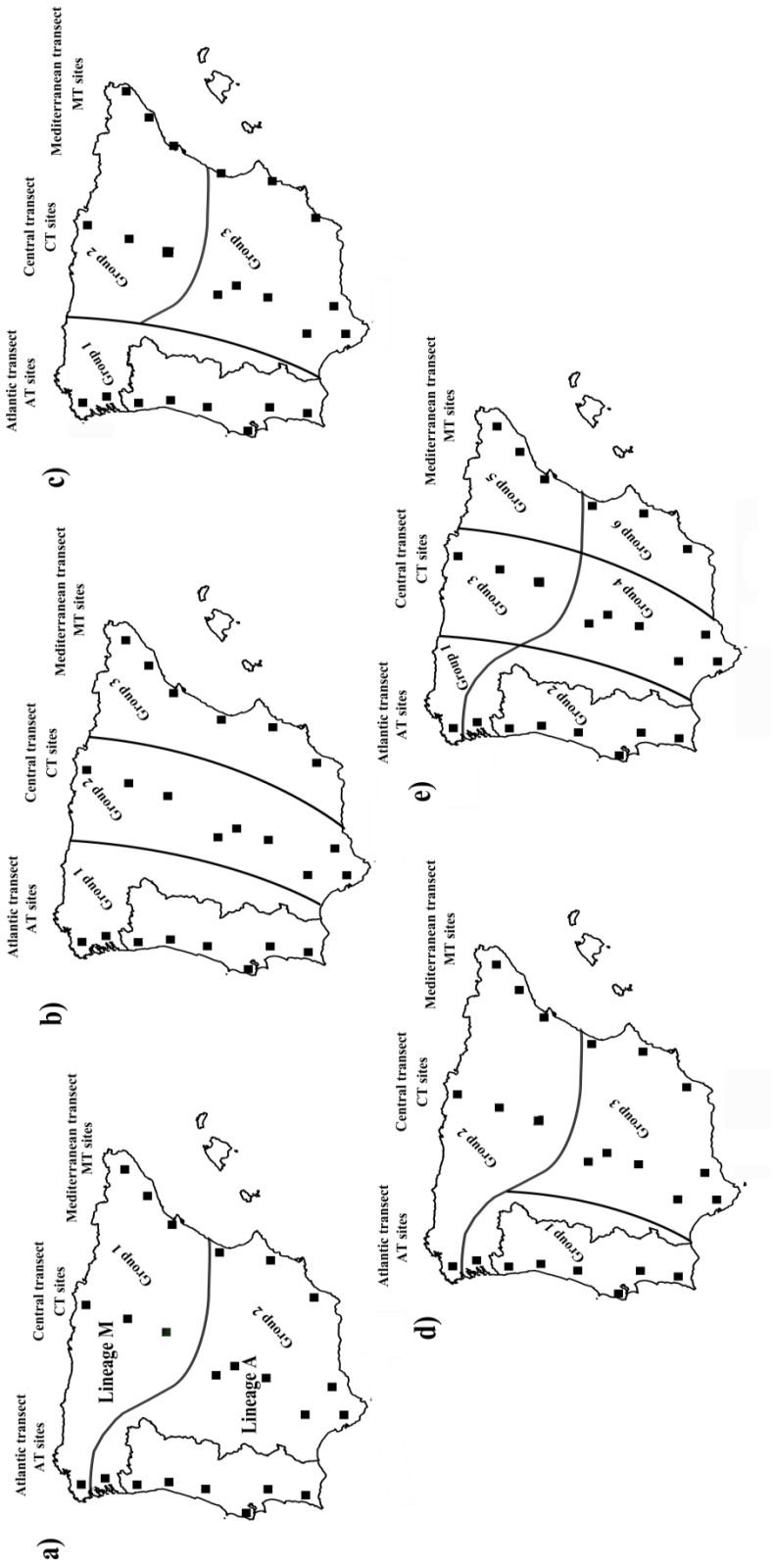


Figure Sup V - 2. Graphical representation of the five groupings used in the AMOVA (see Table V - 2). Letters "a)" to "e)" in this Figure correspond to the letters "a)" to "e)" in the footnotes of Table V - 2.

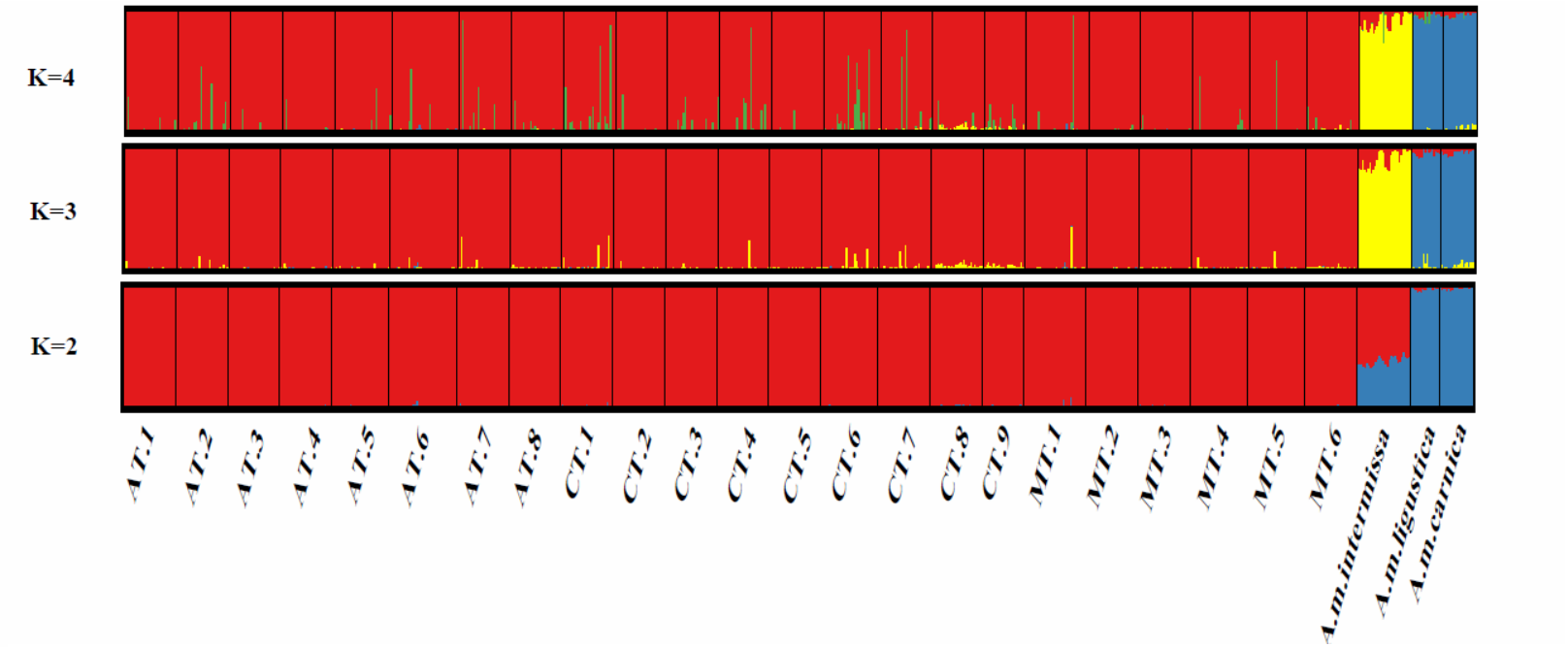


Figure Sup V - 3. Population structure and admixture levels obtained with the software STRUCTURE based on 1075 SNP loci. Each individual is represented by a bar, which is partitioned into K colored segments that represent the individual's estimated membership proportions in K clusters. Black lines separate individuals of the focal subspecies *A. m. iberiensis* (AT1 to MT6) and of the reference subspecies *A. m. intermissa, A. m. ligustica*, and *A. m. carnica*. The number of ancestral clusters (K) was estimated using the admixture ancestry and correlated allele frequency models with the unsupervised option. STRUCTURE was set up for 750,000 Markov chain Monte Carlo iterations after an initial burn-in of 250,000. Over 20 independent runs for each K (from 2 to 4) were performed to confirm consistency across runs. The Greedy algorithm, implemented in the software CLUMPP 1.1.2 (Jakobsson & Rosenberg 2007), was used to compute the pairwise "symmetric similarity coefficient" between pairs of runs and to align the 20 runs for each K. The means of the permuted results were plotted using the software DISTRUCT 1.1 (Rosenberg 2004).

References

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**, 1801-1806.

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes* **4**, 137-138.

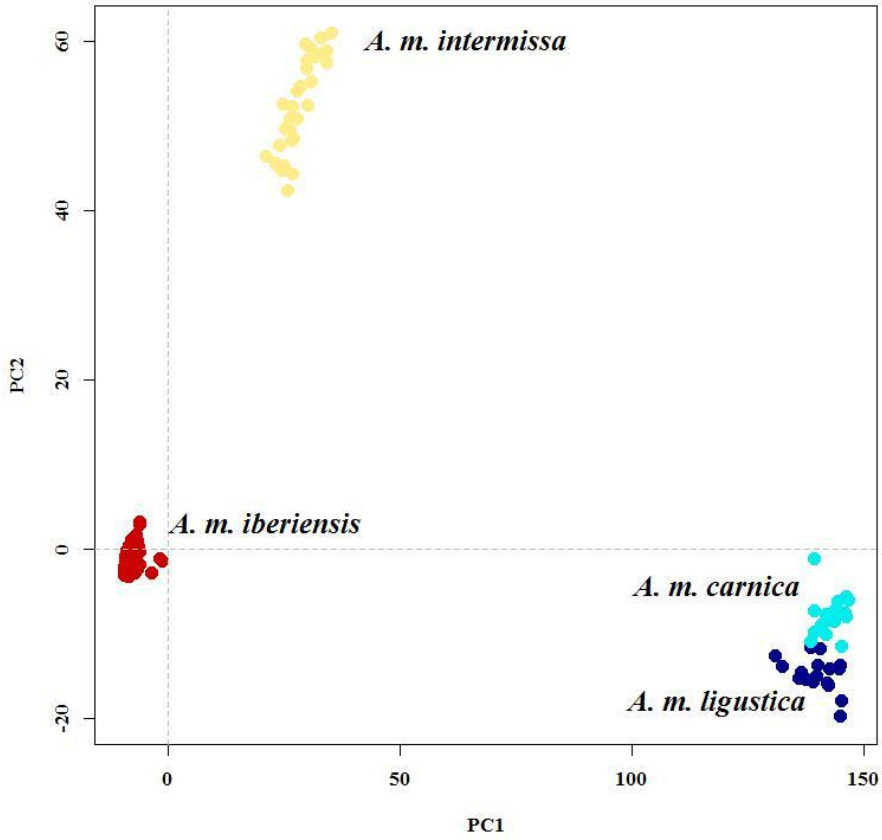


Figure Sup V - 4. Principal component analysis (PCA) based on 1075 SNP loci. PCA was performed on a normalized matrix of individuals *versus* SNP loci. As PCA is sensitive to missing data, genotypes were imputed for missing values using the mean allele frequency through the function ScaleGen available in ADEGENET. Principal components and variances were calculated from the singular value decomposition. PC1 separates *A. m. iberiensis* from the two eastern European beekeepers-favorite subspecies *A. m. ligustica* and *A. m. carnica* whereas PC2 separates the northern African *A. m. intermissa*. PCA 1 and PCA2 explain 48.1 % and 5.6 % of the variance, respectively.

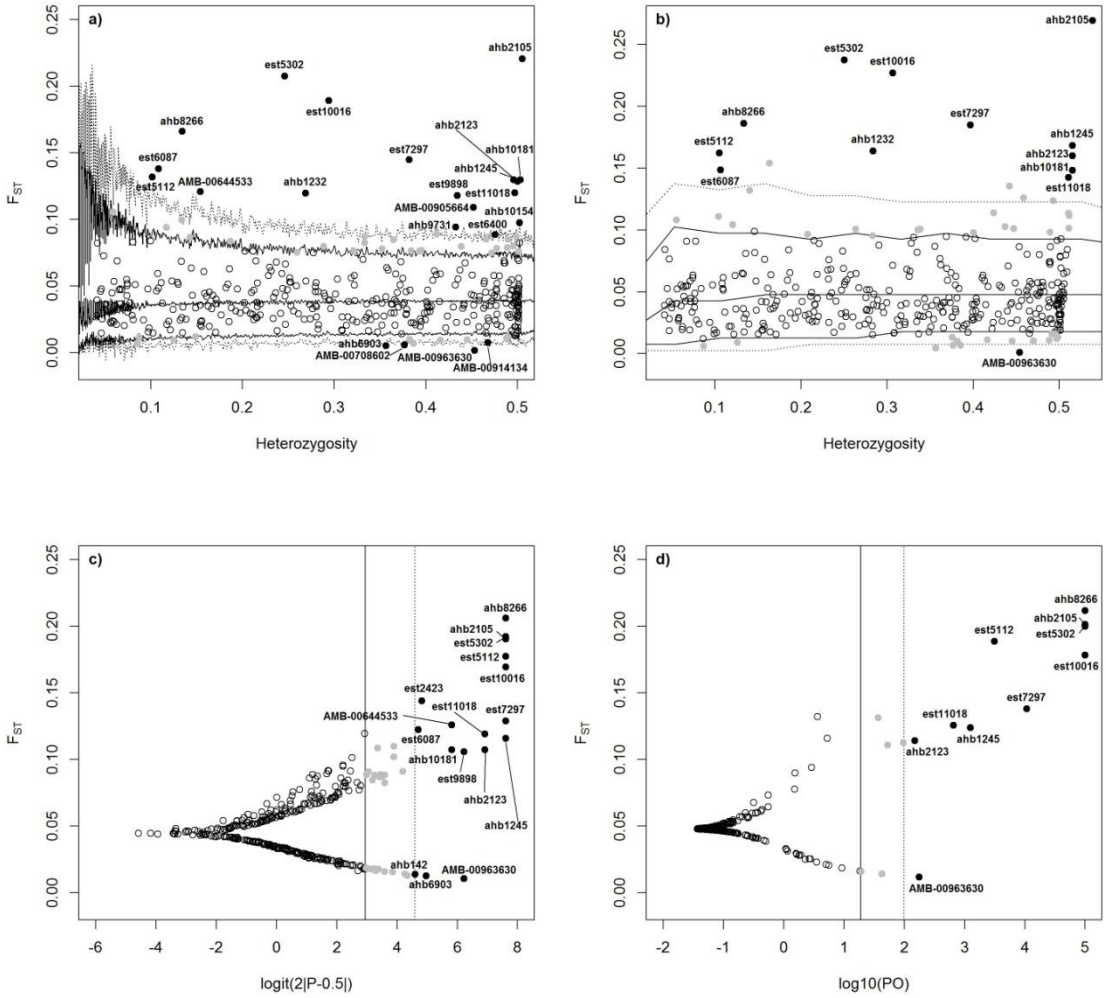


Figure Sup V - 5. a) LOSITAN, b) ARLEQUIN, c) BAYESFST, and d) BAYESCAN plots showing loci under selection. Solid and dashed lines represent 95 % and 99 % confidence intervals, respectively. The middle line in a) and b) depicts the median value. Outlier directional (upper) and balancing (lower) SNP loci with *P-value* ≤ 0.005 and posterior probability ≥ 0.99 and are labeled in each plot.

Table Sup V - 1. Genomic information obtained from NCBI (http://www.ncbi.nlm.nih.gov/), BEEBASE (http://hymenopteragenome.org/beebase), and FLYBASE (www.flybase.org) for the remaining directional and balancing outlier SNP loci as detected by two and one F_{ST}-based method (*P*values ≤ 0.05), and for SNP loci that were associated with at least one environmental variable identified by the spatial analysis method (matSAM).

SNP code	Linkage group	Position	Gene product	Putative function	SNP location	Amino acid change	BAYESFST (*P*value)	ARLEQUIN (*P*value)	LOSITAN (*P*value)	SAM (CI≥ 99%)
Loci detected by two F_{ST}-based methods										
Directional										
AMB-00331869	2	7238233	Hypothetical protein LOC100577383	Unknown	Intron			0.0113	0.0086	Long, Ins
est6400	8	10146419	Hypothetical protein LOC725263 (DOR)[s]	Regulation	3'-UTR			0.0096	0.0039	Long, Prec
ahb10301	9	3114028	Hypothetical protein LOC725264 (beaten path IIa, beat-IIa)[s]	Signaling	Intron			0.0397	0.0231	
ahb2293	11	1001885	Hypothetical protein LOC100578739	Unknown	Intron			0.0427	0.0277	Long
AMB-01023229	11	8108936	UNC93 - like protein	Transport	Intron		0.0170	0.0294		Long
est9912	13	3129509	AP 50 - Clathrin coat adaptor protein isoform1 (Adaptor protein complex 2 μ subunit, AP-2μ)[s]	Transport	Exon	Gly/Gly		0.0204	0.0166	Long, Prec, Cld
AMB-01003140***	0	167966	Histone-lysine N-methyltransferase SETMAR-like, pseudogene	Unknown	11267			0.0406	0.0206	Prec
Balancing										
AMB-00931685**	4	7390795						0.0207	0.0094	
ahb7354***	5	3663050	Down syndrome cell adhesion molecule homolog isoform 1	Unknown	25072		0.0165		0.0225	
AMB-00731725***	5	7937240	Homeobox protein homothorax-like (homothorax, hth)[s]	Regulation	47960		0.0245	0.0358		

Loci detected by one F_{ST}-based method										
Directional										
AMB-00336076*	2	3543913	Golgi phosphoprotein3-like/ Lachesin-like	Regulation/Structural	4424/16338				0.0269	
ahb5992***	2	6678143	Hypothetical protein LOC100577125	Unknown	16117			0.0493		
ahb5536**	2	8020776					0.0225			
AMB-00707983	2	9472652	Epidermal growth factor receptor kinase substrate 8-like protein 1-like (arouser, aru)[s]	Regulation	Exon	Ala/Ala			0.0202	Cld
est1808	2	10050307	Saccharopine dehydrogenase-like isoform 1	Metabolism	Intron			0.0363		
est2200	2	10959133	Solute carrier family 41 member 2 – like protein	Transport	3'-UTR				0.0094	Prec, Rh
est3866	4	4017084	Alanine-glyoxylate aminotransferase 2 - like	Metabolism	Exon	Val/Val		0.0449		
AMB-01082776	5	1210156	Hypothetical protein LOC411058	Unknown	Intron				0.0187	Ins
est4488	5	12883524	J domain-containing protein-like isoform 2 (jdp)[s]	Structural	3'-UTR				0.0129	
ahb8709	6	12351624	Hypothetical protein LOC100577246	Unknown	Intron				0.0222	
ahb8763est5330	6	15417455	Glucocerebrosidase isoform 1	Metabolism	Exon	Gln/Gln			0.0078	Prec
est6729	9	5040827	Long-chain-fatty-acid-CoA ligase 3-like isoform 1 (Acyl-CoA synthetase long-chain, Acsl)[s]	Regulation	Exon	Asp/Asp			0.0093	
est8308	11	1444579	Leucine-rich repeat-containing protein C10orf11 homolog	Unknown	Exon	Asn/Asn			0.0241	
est8691*	11	12638059	Ubiquitin-like-conjugating enzyme ATG10-like isoform 1/Ssdp - sequence-specific single-stranded DNA-binding protein isoform 2 (Sequence-specific single-stranded DNA-binding protein, Ssdp)[s]	Signaling/Regulation	5'-UTR/5'-UTR		0.0100			Long

est9606	12	8158899	Porin - Voltage-dependent anion-selective channel (porin)[s]	Transport	Exon	Asn/Asn			0.0173	Prec, Cld
est9938	13	3063549	RING finger protein nhl – 1-like (thin tn)[s]	Structural	Exon	Val/Val			0.0194	
est10068	13	4537965	Probable 4-coumarate-CoA ligase 3-like (pudgy, pdgy)[s]	Metabolism	Exon	Gln/Arg	0.0185			Prec, Cld
est11407	15	8294367	RpL26 – Ribosomal protein L26 isoform 1 (Ribosomal protein L26, RpL26)[s]	Structural	Exon	His/His			0.0208	
est11786	16	5705216	Hypothetical protein LOC408561	Unknown	5'-UTR				0.0256	
ahb5249*	16	6841443	acj6 - abnormal chemosensory jump 6 (abnormal chemosensory jump 6, acj6)[s]/P protein-like isoform 2 (hoepel1, hoe1)[s]	Regulation/Transport	5703/7766				0.0061	
Balancing										
est1222	1	3349775	Clk – Clock (Clock, Clk)[s]	Regulation	Exon	Ser/Ser		0.0444		
AMB-01090245**	2	1439361					0.0210			
AMB-00284342	6	4408420	Hypothetical protein LOC550870 isoform 2	Unknown	Intron			0.0282		
AMB-01114739	6	7091877	Hypothetical protein LOC100578775	Unknown	Intron				0.0121	
ahb11439*	16	3215272	Abd-A - Homeobox protein abdominal-A isoform 1 (abdominal A, abd-A)[s]/Hypothetical protein LOC100578791	Regulation/Unknown	17019/43998			0.0328		
Environmental										
ahb6844***	4	7415685	Ras-GEF domain-containing family member 1B-like	Regulation	29526					Long
AMB-00060184	7	1612970	Hypothetical protein LOC411586 (pickled eggs, pigs)[s]	Regulation	Intron					Prec
AMB-00085080	9	3734353	Fasciclin-1 (Fasciclin 1, Fas1)[s]	Signaling	Intron					Long

ahb2383*	11	7862469	Hypothetical protein LOC100578889/ hypothetical protein LOC100578926	Unknown/Unknown	Intron/Intron	Long
ahb5267	16	7139563	Lachesin - like	Structural	Intron	Long

*SNP located between two genes (or putative genes). Function of both genes and physical distances (bp) to the 3' or 5' ends are indicated
**SNP located far away from genes (physical distance varies between 51.8 to 612.6 kb).
***SNP located close to a gene. Physical distance (bp) to the 3' or 5' end of the gene is indicated in the column "SNP location"
§Names and/or symbols within parentheses correspond to orthologous genes of *Drosophila melanogaster* as in FLYBASE
Lat – latitude; Long – longitude; Prec – precipitation; Tmean – mean temperature; Ins – insolation; Cld – cloud cover; Rh – relative humidity
Ala – Alanine; Arg – Arginine; Asn – Asparagine; Asp – Aspartic acid; Gln – Glutamine; Gly - Glycine; His – Histidine; Val – Valine; Ser – Serine

Table Sup V - 2. Environmental variables associated with SNP loci (CI ≥ 99% with Bonferroni correction), as detected by the spatial analysis method (matSAM). Variables are annual, monthly and seasonal. Altitude and land cover are not included in this table as they were not significantly associated with any locus after Bonferroni correction.

Variable	ahb1245	est2423	AMB-00331869	AMB-00190928	AMB-00707983	est2200	ahb6844	AMB-01082776	est5302	ahb8266	ahb8763est5330	AMB-00060184	est5553	ahb9731	est6400	AMB-00085080	ahb10154
Longitude	99.999999999		99.9				99.99						99.99		99.9999999	99.9	
Latitude				99.99					99.999	99.9999999				99.99			99.99999
Precipitation																	
January	99.99					99.9			99.99						99.9999		
February	99.9					99.9					99.9				99.999		
March	99.9					99.9									99.999		
April																	
June				99.99						99.99999				99.99			99.999
August				99.999										99.99			99.99
September				99.99						99.9999				99.99			99.99
October									99.9999			99					
November	99.9					99.9			99.99						99.999		
December	99.99					99.9			99.9						99.9999		
Annual									99.999								
Winter	99.9					99.9									99.999		
Summer				99.999										99.999			99.99
Autumn						99.9			99.999						99.99		
Minimum temperature																	
March	99.9																
April																	
May																	
October																	
November	99																
December	99.9																
Annual																	
Autumn																	
Mean temperature																	
January																	
February																	
March																	
June									99.9	99.999							99.9
November																	
December																	
Winter																	
Autumn																	
Maximum temperature																	
January										99.99							
July									99.99								
August									99.99								
December										99.99							
Cloud cover																	
January									99.999999								
February									99.999999								
March									99.999								
April					99.9				99.9999								

May						99.999999			
June									
July									
August									
September						99.999			
October						99.999999	99.99		
November						99.999999	99.99		
December						99.9999999	99.9999		
Annual						99.9999999			
Winter									
Spring						99.999999			
Summer									
Autumn						99.9999999	99.9999		
Relative humidity									
January									
February									
March									
April									
May									
June									
July									
August									
September									
October									
November									
December									
Annual									
Winter									
Spring									
Summer									
Autumn					99				
Insolation									
January							99.9999999	99.999	99.999
May	99.9		99.9			99.99		99.9999	
June			99.9	99.9				99.9999	99.9
July							99.99	99.999	99.99
August			99.9	99.9			99.9	99.9999	99.99
September							99.9	99.99999	99.99
Spring		99.9	99	99.9		99.9	99.9	99.99999	99.9
Summer							99.9	99.9999	99.99

Table Sup V - 2. (Cont.)

Variable	ahb2123	est7297	ahb2105	ahb2293	ahb2383	AMB-00644533	AMB-01023229	est8691	est9606	est9898	est9912	est10068	est10016	ahb5267	AMB-00119087	AMB-01003140
Longitude	99.9999999	99.99999999	99.9999999999	99.99	99.9	99.9999999	99.999	99.9			99.99999			99.9	99.9	
Latitude	99.9	99.999								99.999999			99.99999			
Precipitation																
January			99.9999999999						99.9		99.999	99.9				
February			99.9999999999						99.9		99.99	99.9				
March			99.99999999						99.9		99.9	99.99				
April			99.9999999						99.9			99.99				
June	99.9	99.99								99.9999			99.999			
August	99.99999	99.9999999				99.99				99.99999					99.99	
September	99.99	99.99999								99.99999			99.999			
October			99.9999						99.9			99.9				99.9
November			99.99999999						99.99		99.99	99.99				
December			99.9999999999						99.9		99.999	99.99				
Annual			99.9999						99.9			99.99				
Winter			99.9999999999						99.9		99.99	99.99				
Summer	99.9999	99.99999								99.99999						
Autumn			99.999999999						99.9		99.99	99.99				
Minimum temperature																
March																
April																
May																
October																
November																
December																
Annual																
Autumn																
Mean temperature																
January																
February																
March																
June										99.9			99.999			
November																
December																
Winter																
Autumn																
Maximum temperature																
January																
July													99.9999999			
August										99.9			99.99999999			
December																
Cloud cover																
January								99.99			99.99	99.999	99.99			
February								99.99			99.999	99.999	99.99			
March								99.99			99.99	99.99				

April				99.99	99.99	99.99
May				99.99	99.99	99.9
June				99	99.99	
July						
August						
September						
October				99.99	99.9	99.9999
November				99.99		99.999
December				99.999		99.999
Annual				99.999	99.9	99.999
Winter				99.99	99.99	99.999
Spring				99.99	99.99	
Summer						
Autumn				99.999		99.9999
Relative humidity						
January	99.99					
February	99.9999					
March						
April						
May						
June						
July						
August						
September						
October		99.9				
November		99.9999				
December		99.9				
Annual						
Winter		99.99				
Spring						
Summer						
Autumn		99.99999				
Insolation						
January					99.9999	99.999
May	99.999	99.99		99.99999	99.99	
June	99.99	99.99		99.999	99.9999	99.9
July		99.99			99.9999	99.999999
August	99.9999	99.9999999		99.9	99.99999	99.9999
September	99.99	99.99999			99.9999	99.99
Spring	99.999	99.99		99.999	99.9999	
Summer	99.99	99.99999			99.99999	99.9999

Table Sup V - 3. Accession numbers of genes marked by SNPs that were detected by at least one of the five methods employed listed in Tables V - 1 and Sup V - 1. Loci exhibiting the strongest signal of selection (see Table V - 1) are marked in bold.

SNP code	BEEBASE	FLYBASE	NCBI
Directional			
AMB-00905664	GB14784	-	XP_003251590.1
ahb1245	GB10566	-	XP_001121652.1
ahb1232	GB10373/GB10459	FBpp0075193/FBpp0083391	XP_624502.3/XP_394264.3
ahb226	-	-	-
est2423	GB11195	FBpp0307929	XP_397308.1
AMB-00336076	GB13437/GB11729	FBpp0100141/FBpp0079377	XP_003249737.1/XP_392251.4
ahb5992	GL630203.1	-	NC_007071.3
AMB-00331869	GL630215.1	-	NC_007071.3
AMB-00190928	GB10381	FBpp0071069	XP_395791.3
ahb5536			
AMB-00707983	GB11269	FBpp0077696	XP_392229.4
est1808	GB17377	FBpp0078472	XP_624456.2
est2200	GB12081	FBpp0112364	XP_394338.2
AMB-00402575	GB16865	FBpp0290709	NP_001229503.1
est3866	GB13140	FBpp0075543	XP_392348.2
AMB-01082776	GB18506	FBpp0070333	XP_394531.3
est4488	GB18301	FBpp0085044	XP_392393.1
est5302	GB13133	FBpp0081851	XP_001120919.1
est5112	GB19405	FBpp0085484	NP_001011570.1
ahb8709	GB11946	-	XP_003251404.1
ahb8266	GB12816	FBpp0078161	XP_394629.4

ahb8763est5330	GB10584	FBpp0306807	XP_393207.2
est5553	GB12741	FBpp0304847	XP_623084.1
est6087	GB11059	FBpp0306689	XP_392490.4
ahb9731	GB15214	-	XP_394746.3
est6400	GB16701	FBpp0300395	XP_001121135.1
ahb10301	GB15181	FBpp0082894	XP_001121136.2
est6729	GB16184	FBpp0087780	XP_393022.3
ahb10181	GB15696/GB12690	-/-	XP_001122470.2/XP_003249971.1
ahb10154	GB12690	-	XP_003249971.1
ahb2123	GB15647	FBpp0305743	XP_396952.3
est7297	GB11685	FBpp0075185	XP_393337.4
ahb2105	GB11481	-	XP_003250178.1
ahb2293	-	-	XP_003249212.1
est8308	GB11492	-	XP_001121245.2
AMB-00644533	GB15081	FBpp0288528	XP_001120126.2
AMB-01023229	GB12202	FBpp0074119	XP_396585.2
est8691	GB13253/GB17557	-/FBpp0303084	XP_001122503.2/XP_623511.2
est9606	GB12113	FBpp0304343	XP_623725.1
est9898	GB13161	FBpp0307859	XP_396711.3
est9938	GB19169	FBpp0085757	XP_391967.4
est9912	GB11701	FBpp0291080	XP_391965.1
est10068	GB19085	FBpp0300995	XP_001121814.2
est10016	GB18052	FBpp0302888	XP_396387.2
est11018	GB13979	FBpp0080166	XP_394098.4

est11407	GB13731	FBpp0305858	XP_392059.1
est11786	GB18669	-	XP_397509.4
ahb5249	GB18833/GB14978	FBpp0073832/FBpp0077088	XP_394134.2/XP_624260.3
AMB-00119087	-	-	-
AMB-01003140	GL630350.1	-	NW_003378187.1
Balancing			
est1222	GB17107	FBpp0076500	XP_394233.4
ahb1129	GB12299	FBpp0293549	XP_397003.4
ahb142	GB11468	FBpp0292882	XP_394163.3
AMB-01090245	-	-	-
AMB-00931685	-	-	-
ahb6903	GB15141	FBpp0110391	NP_001014991.1
ahb7354	GB18796	FBpp0292351	XP_396311.4
AMB-00731725	GB18348	FBpp0081733	XP_624460.3
AMB-00284342	GB18318	FBpp0305357	XP_623264.3
AMB-01114739	-	-	XP_003250413.1
ahb11439	GB19738/-	FBpp0082829/-	XP_394120.4/XP_003249501.1
est5796	GL630350.1	-	NC_007076.3
est6265	GB12034	FBpp0292089	XP_394738.2
AMB-00338679	GB10175	FBpp0302006	XP_623414.3
AMB-00310216	GB10731	FBpp0303177	NP_001191178.1
AMB-00963630	GB12195/GB15920	-/-	XP_003249274.1/XP_003249275.1
ahb4188	GB13493	FBpp0079875	NP_001011606.1
AMB-00708602	GB14452	FBpp0111566	XP_001122665.2

AMB-00914134	GB14359	FBpp0071422	XP_001123207.2
Environmental			
ahb6844	-	-	XP_003250148.1
AMB-00060184	GB16235	FBpp0300876	XP_395056.4
AMB-00085080	GB15085	FBpp0082783	XP_394818.4
ahb2383	-/-	-/-	XP_003250460.1/XP_003250459.1
ahb5267	GB11316	FBpp0292059	XP_394132.4

*Accession number of DNA sequence

Supporting information: Chapter VI

Table Sup VI - 1. Statistics of physical distances (bp) of the 309 neutral SNPs used in the genetic analysis of the Iberian honey bee.

Linkage group	Number of SNPs	Minimum distance (bp)	Maximum distance (bp)	Average distance (bp)
1	38	9675	3207376	749970.2
2	15	55085	3549486	889187.2
3	17	72953	2063052	713281.4
4	13	60766	2597150	926669.3
5	22	13345	2200186	654613.5
6	29	3760	2191245	622479.3
7	31	1669	1528031	430482.5
8	19	16903	2318348	718235.2
9	11	57985	4509270	1070820
10	12	362819	2611263	1102939
11	16	3233	2085857	857835.5
12	16	71555	1229619	617181.3
13	17	3417	2200956	570972.5
14	15	19995	1199607	622720.3
15	16	85988	1705206	628494.9
16	11	33188	1733799	634971.8
0*	11			

* Unplaced SNPs in the honey bee genome.

Table Sup VI - 2. Pairwise Φ_{PT} values among sampling sites estimated from the neutral SNP dataset with GENALEX 6.5 (Peakall & Smouse 2012). Significance of Φ_{PT} estimates was assessed using 10,000 permutations. Global Φ_{PT} value was 0.020 (*P*value=0.001). Pairwise Φ_{PT} values ranged from 0.000 to 0.046. Φ_{PT} values marked in bold were significantly different from zero following Bonferroni correction. Sampling site codes are specified in Fig. VI – 1.

Sampling sites	AT1	AT2	AT3	AT4	AT5	AT6	AT7	AT8	CT1	CT2	CT3	CT4	CT5	CT6	CT7	CT8	CT9	MT1	MT2	MT3	MT4	MT5	MT6
AT1																							
AT2	0.004																						
AT3	**0.012**	0.011																					
AT4	**0.022**	**0.016**	0.010																				
AT5	**0.030**	**0.024**	**0.018**	**0.014**																			
AT6	**0.022**	**0.021**	**0.019**	**0.016**	0.008																		
AT7	**0.025**	**0.025**	**0.028**	**0.022**	0.008	0.000																	
AT8	**0.034**	**0.032**	**0.025**	**0.024**	**0.016**	0.011	0.006																
CT1	**0.025**	**0.032**	**0.032**	**0.027**	**0.040**	**0.032**	**0.028**	**0.035**															
CT2	**0.031**	**0.035**	**0.032**	**0.031**	**0.034**	**0.028**	**0.022**	**0.033**	0.011														
CT3	**0.033**	**0.030**	**0.045**	**0.037**	**0.042**	**0.039**	**0.032**	**0.040**	**0.023**	**0.029**													
CT4	**0.031**	**0.031**	**0.024**	**0.026**	**0.023**	**0.018**	0.014	**0.022**	**0.023**	**0.027**	**0.031**												
CT5	**0.022**	**0.024**	**0.018**	**0.021**	**0.019**	**0.012**	0.009	**0.019**	**0.023**	**0.021**	**0.031**	0.008											
CT6	**0.032**	**0.025**	**0.023**	**0.027**	**0.023**	**0.017**	**0.017**	**0.022**	**0.022**	**0.023**	**0.026**	0.011	**0.012**										
CT7	**0.022**	**0.018**	**0.015**	**0.016**	0.011	0.010	0.009	**0.014**	**0.016**	**0.019**	**0.025**	0.008	0.008	0.006									
CT8	**0.037**	**0.036**	**0.034**	**0.030**	**0.029**	**0.027**	**0.023**	**0.026**	**0.037**	**0.029**	**0.044**	**0.026**	**0.022**	**0.016**	**0.015**								
CT9	**0.030**	**0.034**	**0.027**	**0.025**	**0.020**	**0.021**	**0.014**	**0.023**	**0.027**	**0.024**	**0.037**	0.016	0.012	**0.015**	0.009	0.008							
MT1	**0.021**	**0.036**	**0.035**	**0.039**	**0.044**	**0.033**	**0.036**	**0.042**	**0.014**	**0.025**	**0.039**	**0.037**	**0.026**	**0.036**	**0.029**	**0.046**	**0.034**						
MT2	**0.029**	**0.039**	**0.035**	**0.039**	**0.045**	**0.043**	**0.039**	**0.041**	**0.016**	**0.018**	**0.038**	**0.038**	**0.023**	**0.037**	**0.028**	**0.046**	**0.039**	0.007					
MT3	**0.021**	**0.026**	**0.022**	**0.026**	**0.025**	**0.021**	**0.020**	**0.022**	0.008	0.009	**0.027**	**0.020**	0.010	**0.015**	**0.012**	**0.028**	**0.022**	0.011	0.005				
MT4	**0.016**	**0.023**	**0.022**	**0.022**	**0.023**	**0.013**	**0.015**	**0.019**	0.011	**0.013**	**0.028**	**0.013**	0.007	0.011	0.008	**0.015**	0.009	**0.013**	**0.012**	0.006			
MT5	**0.025**	**0.032**	**0.026**	**0.026**	**0.023**	**0.020**	**0.017**	**0.018**	**0.016**	**0.013**	**0.026**	**0.014**	0.006	**0.014**	0.004	**0.020**	**0.020**	**0.021**	**0.016**	0.001	0.004		
MT6	**0.021**	**0.031**	**0.023**	**0.024**	**0.023**	**0.016**	**0.012**	**0.016**	0.011	0.014	**0.033**	**0.015**	0.010	**0.014**	0.008	**0.015**	0.010	**0.022**	**0.023**	0.007	0.003	0.006	

Table Sup VI - 3. Cline parameter estimates for the best-fitting model of 33 SNP loci, complete SNP dataset, neutral SNP dataset and mtDNA. Cline width is presented as 1/maximum slope. Cline center and width are measured in km, *P*min and *P*max are the allele frequencies at the ends of the cline, and δ and τ are the shape parameters for the mirror (M), left (L) and right (R) tails. Two log-likelihood unit support limits are presented in parentheses. The symbol · indicates coincident center and the symbol · indicates concordant width, based on LRTs (*P*-value > 0.05). The left side AICc corresponds to the best-fitting model, and the right side AICc corresponds to the null model. Neutral SNP loci are marked in bold.

Locus	Linkage group	Position in LG	Best-fitting model	Center	Width	*P*min	*P*max	δL	τL	δR	τR	δM	τM	AICc
est102	1	12904387	Pmin/Pmax observed, left tail	938.8 (734.9 - 1225.7)	846.9 (147.6 - 1393.9)	0.3	0.7			153.5 (0.0 - 472.8)	0.004 (0.0 - 0.8)			33.0/37.8
AMB-00869828··	1	13446522	Pmin/Pmax observed, no tails	550.5 (403.1 - 703.3)	1393.7 (1063.0 - 1393.9)	0.3	0.7							32.6/45.4
AMB-00905664·	1	15234278	Pmin/Pmax observed, left tail	604.8 (432.5 - 745.3)	899.2 (421.6 - 1393.3)	0.3	0.9	30.71 (0.0 - 158.9)	0.004 (0.0 - 0.4)					45.5/88.8
ahb12048··	1	24839613	Pmin/Pmax observed, no tails	705.9 (517.1 - 893.1)	1393.6 (944.5 - 1394)	0.1	0.4							30.4/37.4
ahb1245··	1	25131116	Pmin/Pmax observed, no tails	776.9 (687.8 - 872.1)	1348.3 (1027.6 - 1393.9)	0.3	0.9							45.8/116.6
ahb226··	1	29654719	Pmin/Pmax observed, right tail	661.8 (518.9 - 792.6)	1329.1 (772.4 - 1393.9)	0.0	0.5			145.5 (0.1 - 397.9)	0.01 (0.00 - 0.9)			38.0/64.5
est1808··	2	10050307	Pmin/Pmax observed, no tails	537.9 (424.2 - 656.6)	1191.2 (793.8 - 1393.9)	0.2	0.7							28.4/72.6
AMB-00402575··	3	5547433	Pmin/Pmax observed, no tails	554.3 (443.1 - 665.8)	1393.6 (1072.3 - 1393.9)	0.0	0.5							41.9/77.7
ahb6599·	3	6978016	Pmin/Pmax observed, left tail	745.6 (540.2 - 806.2)	398.6 (282.1 - 1369.7)	0.3	0.8	4.67 (0.0 - 57.1)	0.01 (0.0 - 0.2)					29.8/45.1
AMB-00815487··	3	7519470	Pmin/Pmax observed, left tail	649.4 (534.0 - 993.7)	1369.7 (328.5 - 1393.9)	0.5	0.9	195.61 (0.0 - 379.5)	0.01 (0.0 - 0.7)					33.8/46.6
AMB-00679056	5	1475134	Pmin/Pmax observed, mirror tails	406.9 (234.1 - 505.9)	458.7 (78.8 - 1393.9)	0.4	0.9					45.2 (0.1 - 548.8)	0.2 (0.0 - 0.9)	41.2/56.2
AMB-00046221·	5	5549316	Pmin/Pmax estimated, no tails	363.2 (218.5 - 491.5)	1393.0 (1058.9 - 1393.9)	0.0	0.4							41.1/48.5
est4575··	5	13447645	Pmin/Pmax observed, no tails	711.9 (569.3 - 842.6)	1393.9 (920.2 - 1393.9)	0.0	0.4							28.5/53.4
AMB-00717792	6	4065134	Pmin/Pmax fixed, mirror tails	463.5 (305.7 - 555.3)	953.9 (73.5 - 1393.9)	0.0	1.0					58.6 (0.1 - 136.7)	0.02 (0.0 - 0.3)	34.2/39.4
est6400··	8	10146419	Pmin/Pmax observed, no tails	661.3 (564.6 - 759.5)	1393.9 (1101.4 - 1393.9)	0.1	0.7							28.5/79.6
AMB-00283205·	8	11444452	Pmin/Pmax observed, left tail	872.9 (714.7 - 1151.1)	1384.4 (244.5 - 1393.9)	0.2	0.8	246.57 (0.1 - 361.3)	0.01 (0.0 - 0.7)					31.2/42.9
AMB-00338679·	8	11725179	Pmin/Pmax observed, no tails	492.7 (252.4 - 747.9)	1392.6 (724.0 - 1393.9)	0.4	0.7							19.1/21.5
est7565··	10	703516	Pmin/Pmax observed, no tails	635.7 (457.3 - 815.8)	1393.8 (986.7 - 1393.9)	0.3	0.6							30.4/36.9
ahb2123·	10	2249429	Pmin/Pmax observed, no tails	786.7 (702.3 - 873.7)	1393.9 (1190.6 - 1393.9)	0.2	0.9							45.2/111.6
est7297··	10	3151832	Pmin/Pmax observed, no tails	769.5 (675.0 - 854.5)	1388.5 (1075.1 - 1393.9)	0.0	0.6							45.8/124.3
ahb2105··	10	6427742	Pmin/Pmax fixed, mirror tails	609.3 (538.1 - 660.9)	1161.0 (893.1 - 1393.9)	0.0	1.0					301.0 (196.7 - 408.5)	0.03 (0.0 - 0.7)	74.9/186.8
ahb2293	11	1001885	Pmin/Pmax estimated, no tails	547.6 (523.3 - 704.5)	41.1 (0.5 - 408.3)	0.2 (0.1 - 0.2)	0.4 (0.3 - 0.5)							24.0/66.8
ahb2383·	11	7862469	Pmin/Pmax observed, no tails	873.0 (730.3 - 1019.9)	1393.2 (909.0 - 1393.9)	0.3	0.8							20.9/49.9
AMB-00776332	12	4447603	Pmin/Pmax observed, left tail	962.6 (761.23 - 1130.2)	964.3 (209.5 - 1393.9)	0.0	0.6	5.51 (0.1 - 414.3)	0.5 (0.0 - 0.9)					28.9/58.4
ahb3142··	12	6836872	Pmin/Pmax estimated, no tails	436.6 (276.2 - 592.2)	1271.5 (644.4 - 1393.9)	0.1	0.4							20.8/36.8
est9669··	13	2702684	Pmin/Pmax observed, right tail	588.9 (233.8 - 726.1)	1360.6 (400.8 - 1393.9)	0.2	0.6			192.6 (0.0 - 458.4)	0.001 (0.0 - 0.9)			28.9/36.7
est9912	13	3129509	Pmin/Pmax observed, right tail	467.7	434.5	0.2	0.7			39.8	0.2			25.5/70.7

				(397.5 - 558.5)	(184.4 - 976.9)					(0.1 - 183.1)	(0.0 - 0.7)			
AMB-01112431[·]	15	679566	Pmin/Pmax observed, mirror tails	595.9 (529.4 - 774.9)	380.2 (44.2 - 1393.4)	0.2	0.7					86.9 (6.5 - 345.9)	0.001 (0.0 - 0.6)	30.7/49.4
est11407	15	8294367	Pmin/Pmax estimated, no tails	1320.6 (735.2 - 1363.9)	788.7 (4.4 - 1393.8)	0.5 (0.4 - 0.5)	0.9 (0.6 - 0.9)							58.5/68.2
ahb5244[·•]	16	2005729	Pmin/Pmax observed, mirror tails	548.1 (408.8 - 756.5)	1382.0 (135.3 - 1393.9)	0.1	0.6					337.7 (0.1 - 552.5)	0.01 (0.0 - 0.9)	26.9/41.6
ahb5182[·•]	16	2767375	Pmin/Pmax observed, no tails	664.1 (522.8 - 807.2)	1393.3 (1055.3 - 1393.9)	0.2	0.7							25.2/42.9
AMB-00119087[·•]	0	30896	Pmin/Pmax observed, left tail	759.7 (659.6 - 963.5)	1369.3 (835.9 - 1393.9)	0.4	1.0	299.5 (0.0 - 582.7)	0.1 (0.0 - 0.9)					46.4/78.1
AMB-00914134[·•]	0	78294	Pmin/Pmax observed, no tails	618.8 (397.5 - 838.7)	1392.5 (981.7 - 1393.9)	0.2	0.5							16.9/17.8
Complete SNPs [·•]			Pmin/Pmax fixed, no tails	714.7 (659.1 - 775.3)	1283.5 (1090.5 - 1393.9)	0.0	1.0							25.2/203.5
Neutral SNPs[·•]			Pmin/Pmax observed, no tails	725.7 (651.4 - 808.2)	1047.6 (801.2 - 1393.4)	0.2	0.9							33.2/137.2
MtDNA[·]			Pmin/Pmax fixed, right tail	706.7 (630.7 - 757.9)	580.9 (417.6 - 737.2)	0.0	1.0			6.1 (0.0 - 53.9)	0.3 (0.1 - 0.6)			255.1/516.3

Table Sup VI - 4. Linkage disequilibrium (LD) and unbiased haploid genetic diversity (*uh*) estimated from the 309 neutral SNPs for the 23 sampling sites in the Iberian Peninsula (see Fig. VI - 1 for location of sampling sites), as measured by r^2 and percentage of pairs of loci showing significant LD with the Fisher exact test [P_{LD} (Fisher)] before Bonferroni correction (only a single pair remained significant after Bonferroni correction). None of the SNP pairs exhibiting significant LD before Bonferroni correction, for which there is genomic information, are physically linked as they are located in different chromosomes.

Transect	r^2	P_{LD} (Fisher)	*uh*
AT1	0.036 (±0.0003)	1.91 (594/31125)	0.295 (±0.010)
AT2	0.035 (±0.0003)	1.90 (526/27730)	0.303 (±0.010)
AT3	0.036 (±0.0003)	2.05 (554/27028)	0.285 (±0.010)
AT4	0.035 (±0.0003)	1.98 (563/28441)	0.295 (±0.010)
AT5	0.031 (±0.0002)	1.70 (524/30876)	0.282 (±0.010)
AT6	0.022 (±0.0001)	2.31 (735/31878)	0.286 (±0.010)
AT7	0.035 (±0.0003)	1.68 (475/28203)	0.290 (±0.010)
AT8	0.034 (±0.0003)	1.49 (459/30841)	0.281 (±0.010)
Mean (± SE)	0.033 (±0.001)	1.88 (±0.090)	0.290 (±0.003)
CT1	0.040 (±0.0003)	2.87 (749/26106)	0.309 (±0.010)
CT2	0.020 (±0.0001)	3.11 (975/31375)	0.298 (±0.010)
CT3	0.041 (±0.0003)	3.28 (948/28920)	0.311 (±0.010)
CT4	0.042 (±0.0003)	3.41 (1036/30381)	0.311 (±0.010)
CT5	0.022 (±0.0001)	2.01 (683/33930)	0.299 (±0.010)
CT6	0.034 (±0.0003)	2.37 (673/28441)	0.307 (±0.009)
CT7	0.036 (±0.0003)	1.98 (583/29403)	0.308 (±0.009)
CT8	0.022 (±0.0001)	1.87 (604/32385)	0.299 (±0.010)
CT9	0.045 (±0.0004)	1.63 (505/30876)	0.301 (±0.010)
Mean (± SE)	0.032 (±0.0004)	2.50 (±0.224)	0.305 (±0.003)
MT1	0.029 (±0.0002)	1.94 (525/27028)	0.300 (±0.010)
MT2	0.019 (±0.0001)	2.01 (640/31878)	0.297 (±0.010)
MT3	0.016 (±0.0000)	1.81 (608/33670)	0.306 (±0.010)
MT4	0.014 (±0.0000)	1.65 (551/33411)	0.310 (±0.009)
MT5	0.034 (±0.0000)	1.87 (606/32385)	0.311 (±0.009)
MT6	0.034 (±0.0000)	1.61 (571/35511)	0.313 (±0.009)
Mean (± SE)	0.015 (±0.0001)	1.81 (±0.065)	0.306 (±0.004)
Gran mean (± SE)	0.028 (±0.0000)	2.10 (±0.114)	0.300 (±0.002)

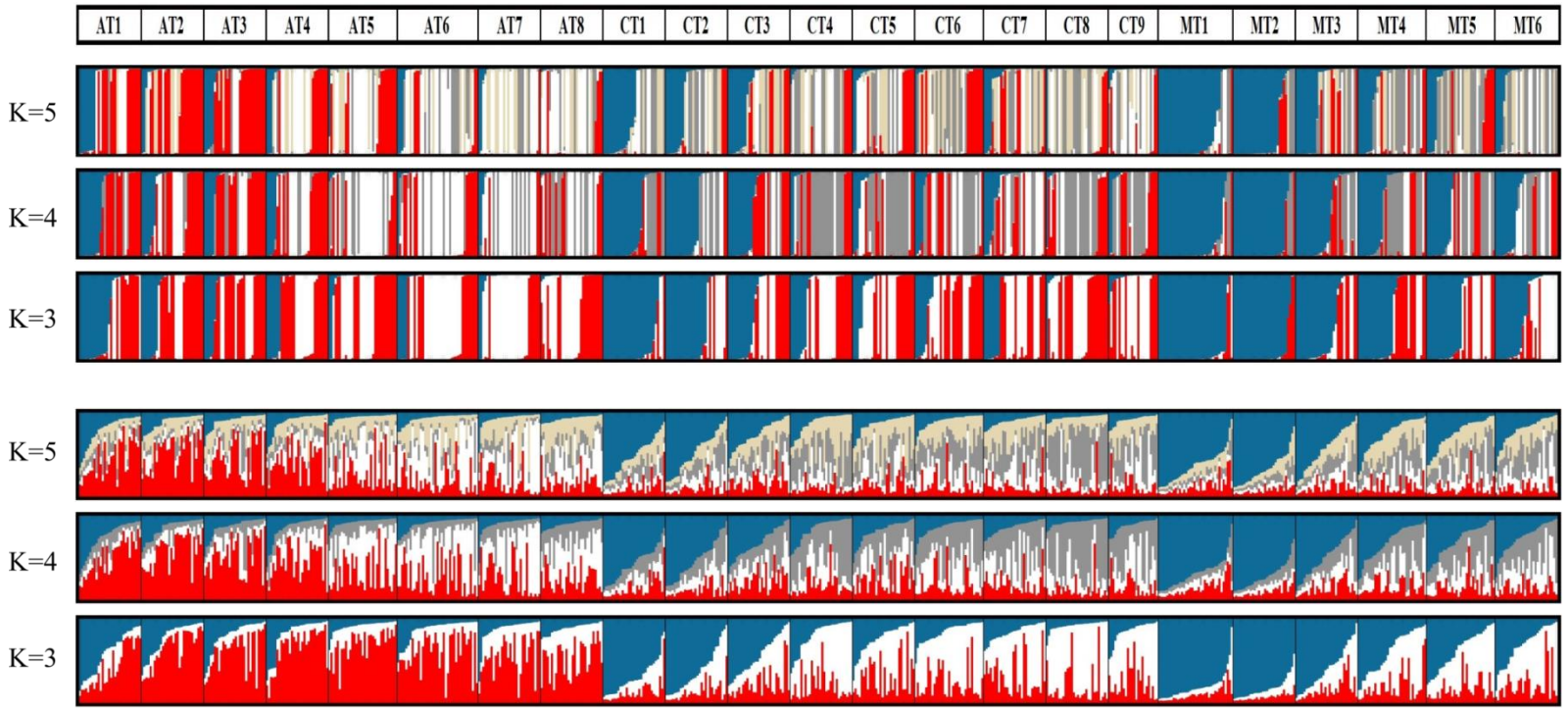


Figure Sup VI - 1. Estimated population structure of *A. m. iberiensis* inferred from the neutral SNP dataset (309 loci) by DAPC (top) and STRUCTURE (bottom) at $K = 3$ to 5 clusters. Each of the 711 individuals included in the analyses is represented by a vertical bar partitioned into colored segments, the size of each corresponding to the individuals' estimated membership proportions in each of the K clusters. Black lines separate individuals from the 23 sampling sites, which are arranged from north (AT1, CT1, MT1) to south (AT8, CT9, MT6) in each of the three transects (AT-Atlantic, CT-central, MT-Mediterranean), as indicated at the top bar. Black lines separate individuals from the 23 sampling sites, which are arranged from high Q (left) to low Q (right) in the blue cluster.

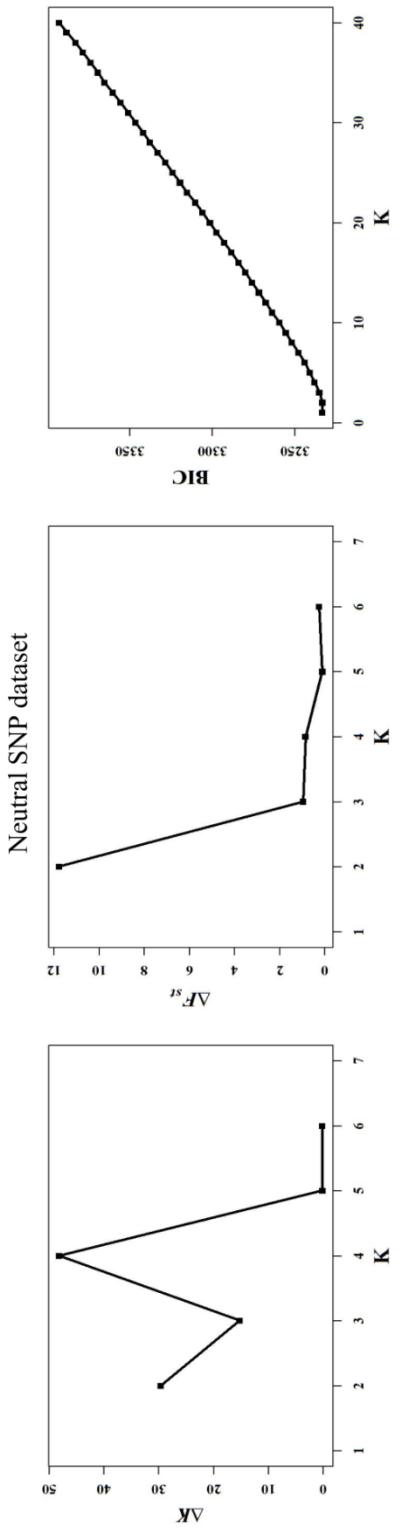


Figure Sup VI - 2. Graphical display of the three methods (ΔK, ΔF_{st} and BIC) to predict the optimal *K* for the analysis of *A. m. iberiensis* population structure using the neutral SNP dataset (309 loci) and the complete SNP dataset (309 neutral plus 74 putatively selected loci identified by Chávez-Galarza *et al.* 2013).

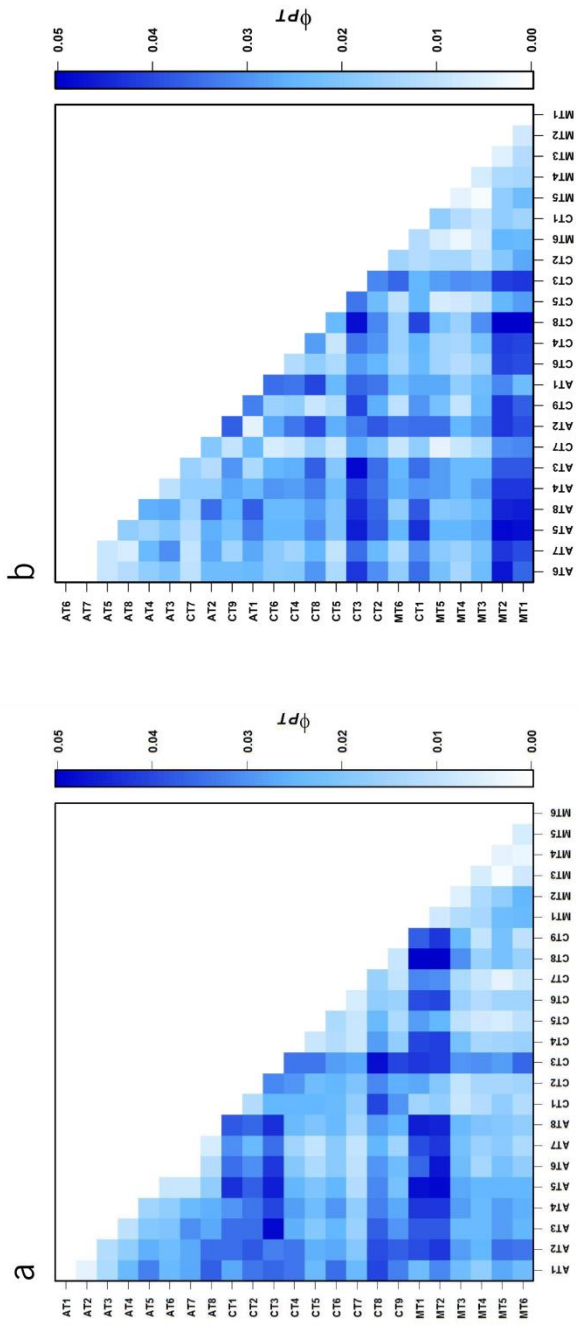


Figure Sup VI - 3. Heat map of pairwise Φ_{PT} values between Iberian sampling sites estimated from the neutral SNP dataset (309 loci) using GENALEX 6.5 (Peakall & Smouse 2012). The heat map clearly highlights northeastern populations (CT1-3, MT1-2) and CT8 as the most differentiated across Iberia. (a) The 23 sampling sites are arranged from north to south in each of the three transects (see Fig. VI - 1). (b) The 23 sampling sites are arranged along the west-east transect traced for the geographic cline analysis (see dashed line in Fig. VI - 6.

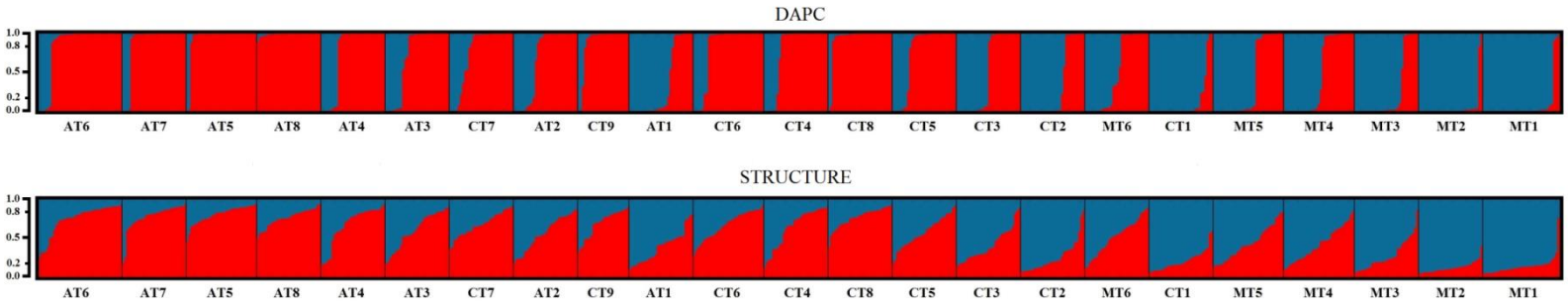


Figure Sup VI - 4. Population structure of *A. m. iberiensis* estimated by (a) DAPC and (b) STRUCTURE from the neutral SNP dataset (309 loci) at $K = 2$ clusters. The 23 sampling sites are arranged along the west-east transect traced for the geographic cline analysis (see dashed line in Fig. VI - 6). Plots represent each of the 711 individuals by a vertical bar partitioned into two colored segments (blue and red) corresponding to membership proportions (Q) in each of the two clusters. Black lines separate individuals from the 23 sampling sites, which are arranged from high Q (left) to low Q (right) in the blue cluster.

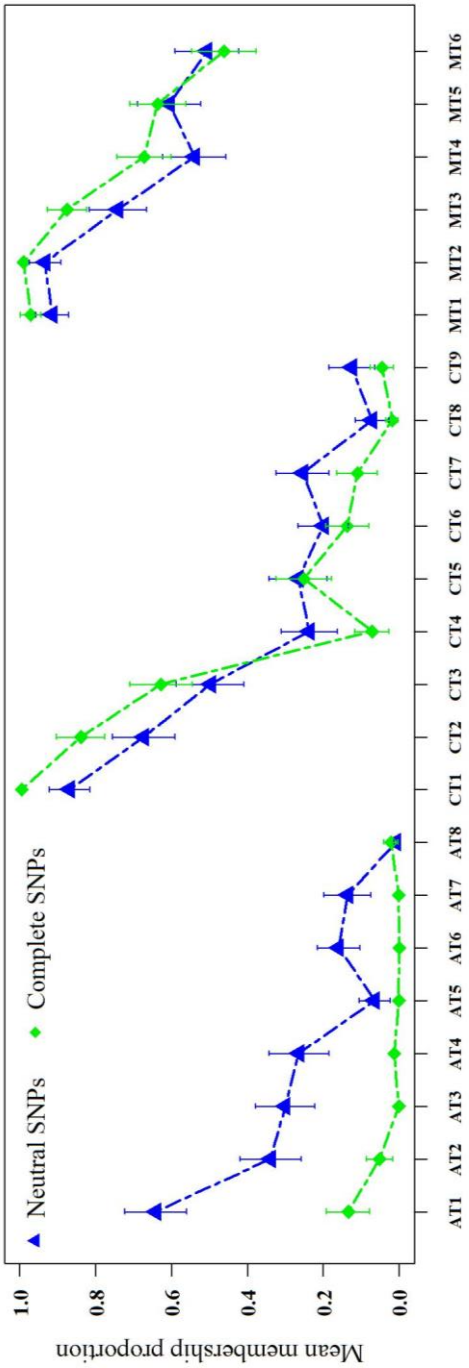


Figure Sup VI - 5. Mean membership proportion (±SE) in the blue cluster inferred from the neutral (309 loci) and the complete SNP dataset (309 neutral plus 74 putatively selected loci identified by Chávez-Galarza *et al.* 2013) with DAPC for each sampling site. Sampling sites are arranged from north (AT1, CT1, MT1) to south (AT8, CT9, MT6) in each of the three transects (AT-Atlantic, CT-central, MT-Mediterranean).

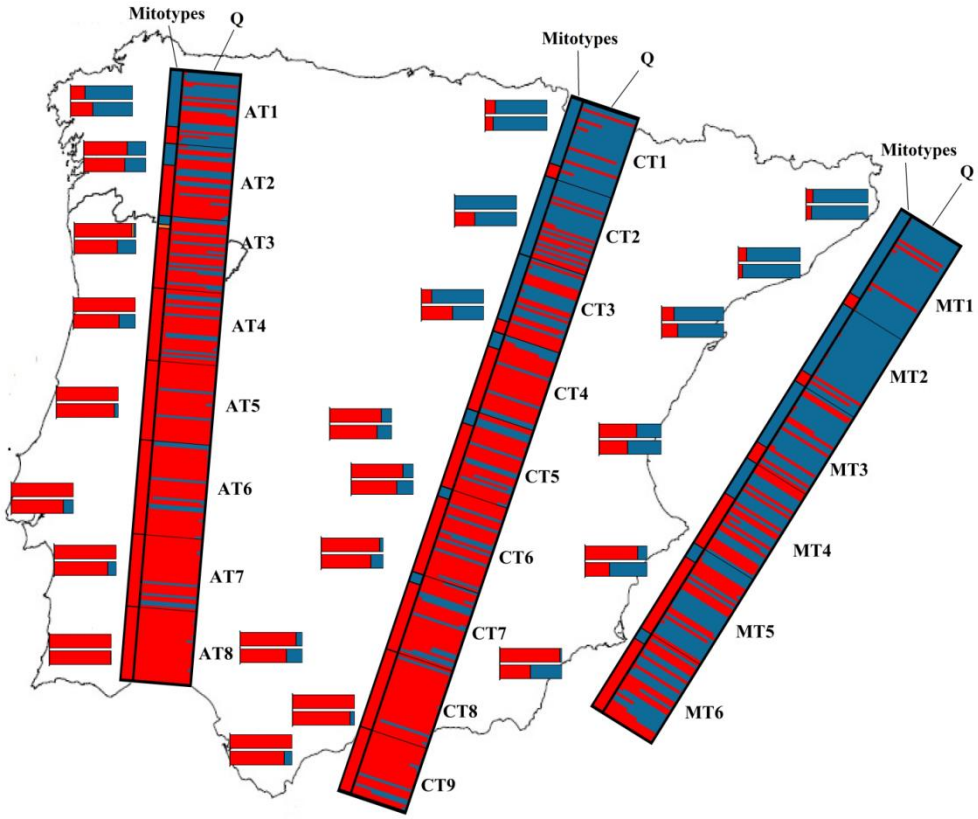


Figure Sup VI - 6. Maternal pattern, obtained from the tRNA^{leu}-cox2 intergenic mitochondrial region, and estimated structure inferred from the neutral SNP dataset (309 loci) with DAPC at $K = 2$ clusters. Patterns of variation are displayed at both individual and sampling site level for each transect (AT-Atlantic, CT-central, MT-Mediterranean). Vertical plots display the mitotype (A in red; M in blue; C in orange, one single individual in AT3) and the membership proportions (Q) for each of the 711 individuals. Horizontal bar plots show mitotype frequencies (top) and the mean Q in blue and red clusters (bottom) at each sampling site.

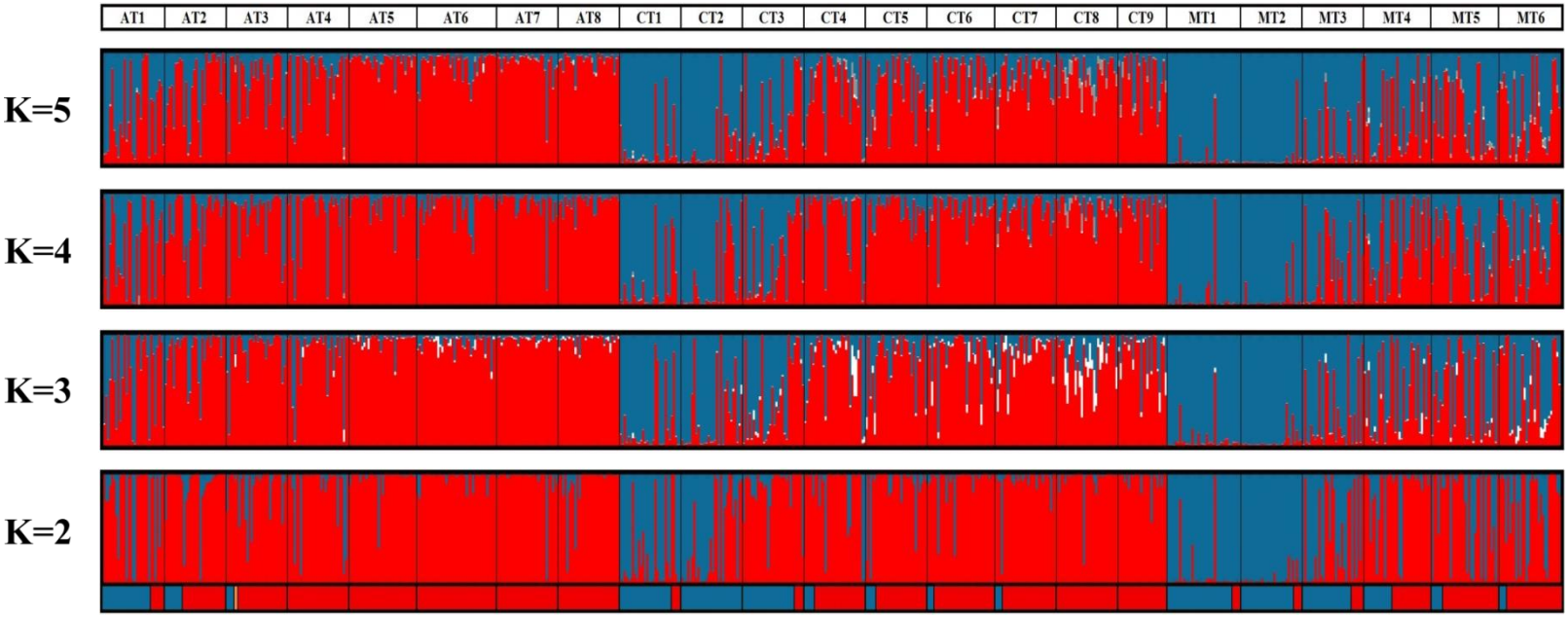


Figure Sup VI - 7. Estimated population structure of *A. m. iberiensis* inferred by the spatially explicit algorithm implemented by TESS for the neutral SNP dataset (309 loci). Each of the 711 individuals included in the analyses is represented by a vertical bar partitioned into colored segments, the size of each corresponding to the individuals' estimated *Q* in each of the *K* (from 2 to 5) clusters. Vertical black lines separate individuals from the 23 sampling sites, which are arranged by M (blue), C (orange, one single individual in AT3), and A (red) maternal lineages indicated by color at the bottom. Sampling sites are arranged from north to south in each transect (AT-Atlantic, CT-central, MT-Mediterranean). Sampling site codes (from AT1 to MT6) are shown at the top bar.

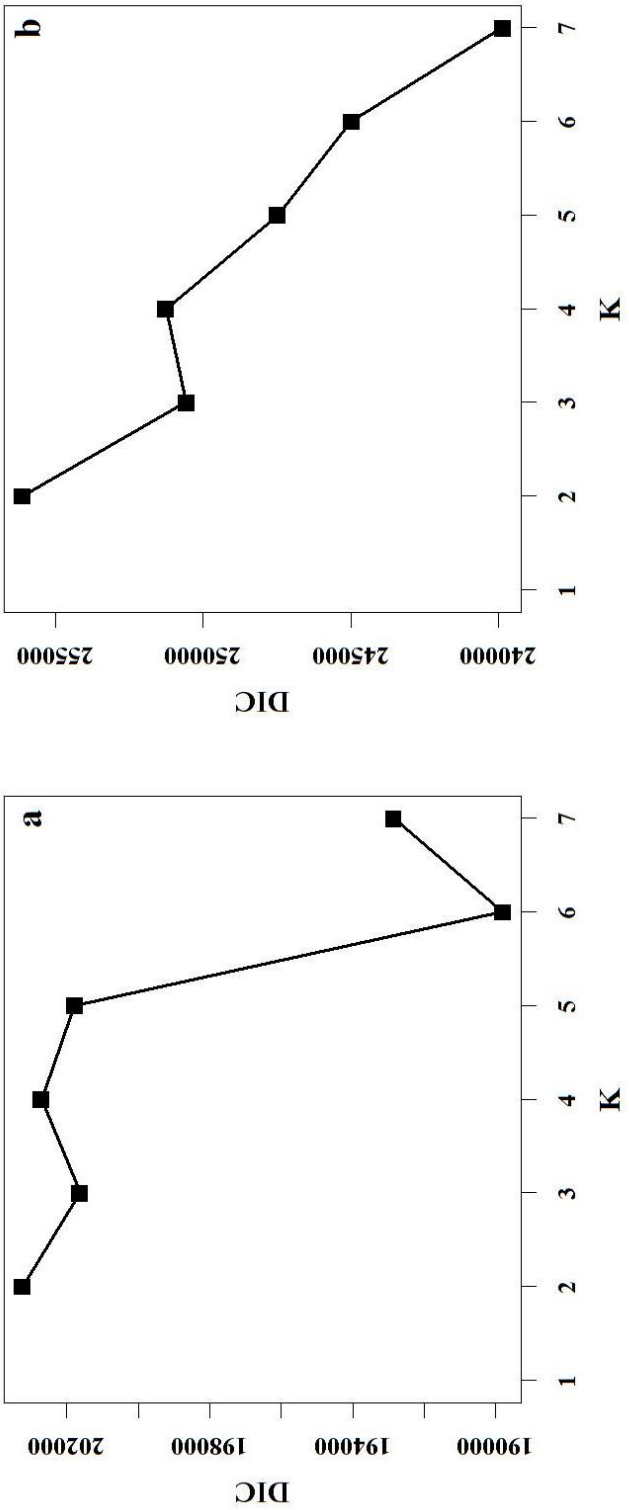


Figure Sup VI - 8. Plot of DIC values (Y-axis) against K_{max} (X-axis) obtained with TESS analysis under the admixture model BYM for (a) the neutral SNP dataset and (b) the complete SNP dataset.

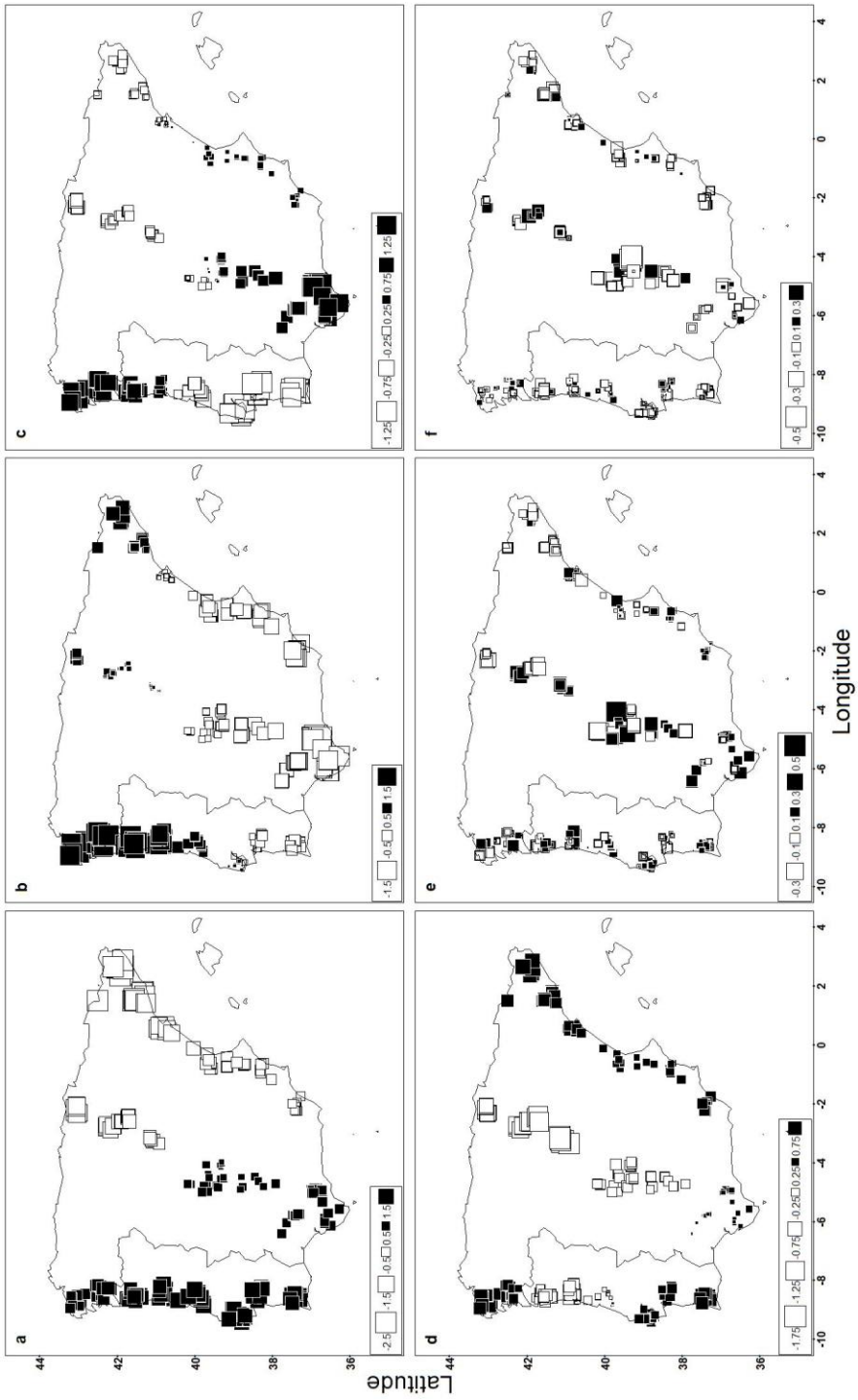


Figure Sup VI - 9. Analysis of global and local genetic structure, among 711 individuals of *A. m. iberiensis* from 23 sampling sites, by spatial principal component analysis (sPCA) using the complete SNP dataset (309 neutral plus 74 putatively selected loci identified by Chávez-Galarza *et al.* 2013). Each square represents the score of an individual, which is positioned by its spatial coordinates. (a-d) The first four global scores of sPCA. (e-f) The first two local scores of sPCA.

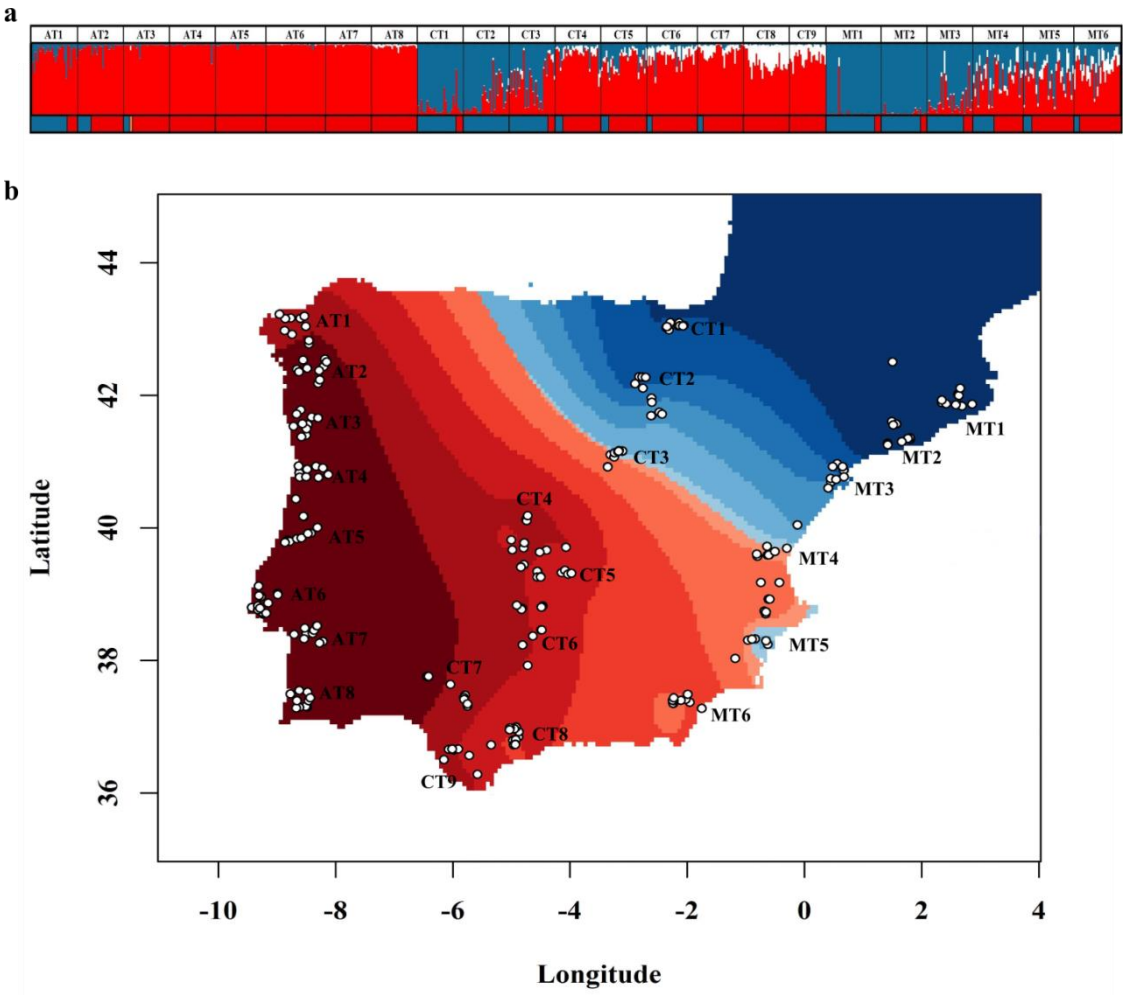


Figure Sup VI - 10. Spatially explicit analysis implemented by the software TESS for the 711 individuals of *A. m. iberiensis* using the complete SNP dataset (309 neutral plus 74 putatively selected loci identified by Chávez-Galarza *et al.* 2013). (a) Plot of individuals' *Q* at the optimal $K = 3$ clusters. Each of the 711 individuals included in the analysis is represented by a vertical bar partitioned into three colored segments (blue, red, and white) corresponding to *Q* in each of the three clusters. Maternal data (M lineage in blue, C lineage in orange, and A lineage in red) are shown at the bottom. Sampling sites and individuals within sampling sites are arranged as in Fig. VI - 3. (b) Map of the Iberian Peninsula showing the two major clusters ($Q \geq 0.5$) interpolated by TESS. Dots represent the locations of sampled apiaries across the Atlantic (AT1-8), central (CT1-9) and Mediterranean (MT1-6) transects.

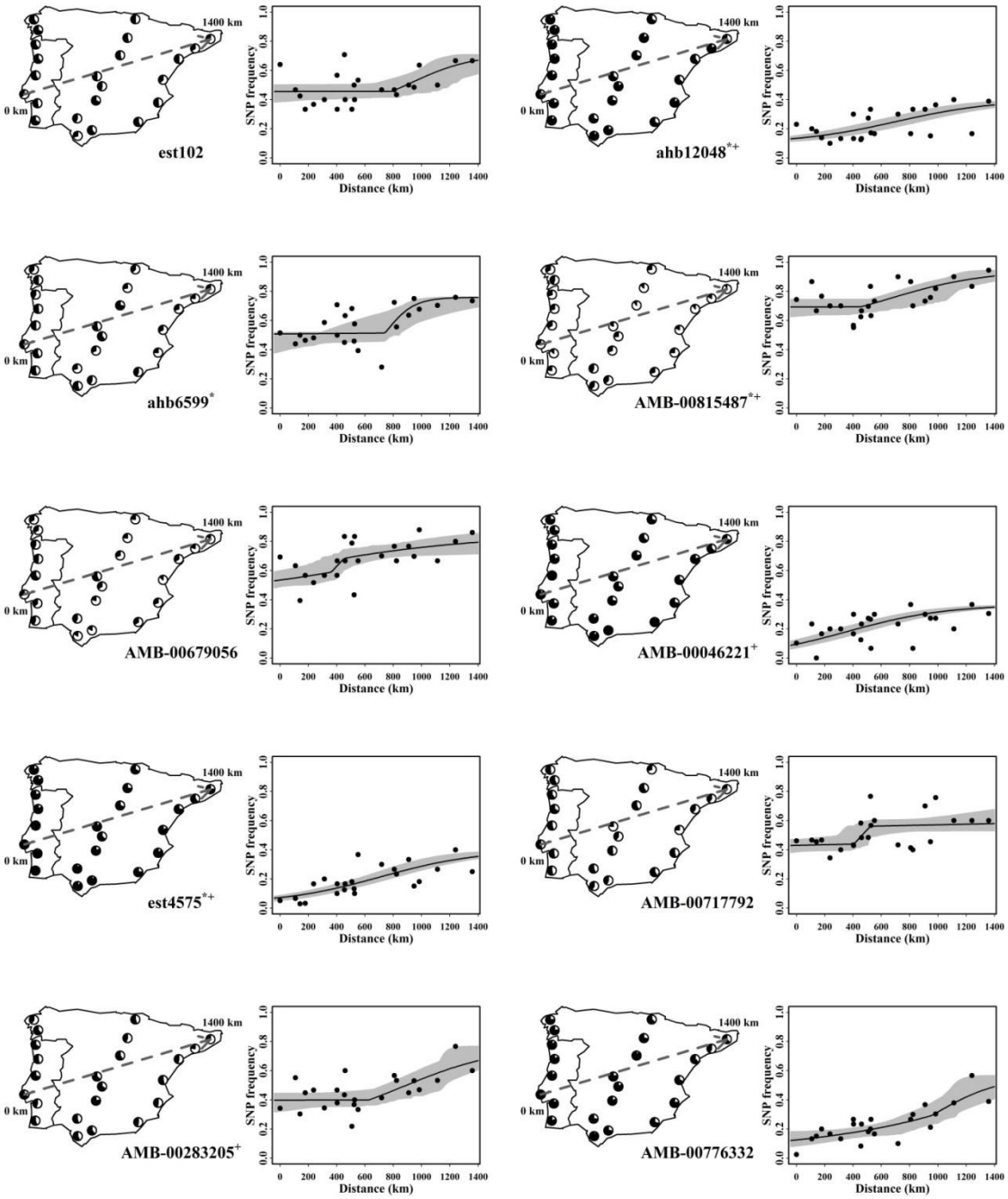


Figure Sup VI - 11. Map of the Iberian Peninsula with pie charts summarizing frequency data for each sampling site and plot of maximum likelihood geographic cline for neutral (marked in bold) and selected SNP loci. The symbols * and + indicate the loci or datasets with coincident center and concordant width, respectively (see Table Sup V - 3).The dashed line placed in each map represents the transect traced from Lisbon (0 km) to Girona (1400 km) for the geographic cline analysis.

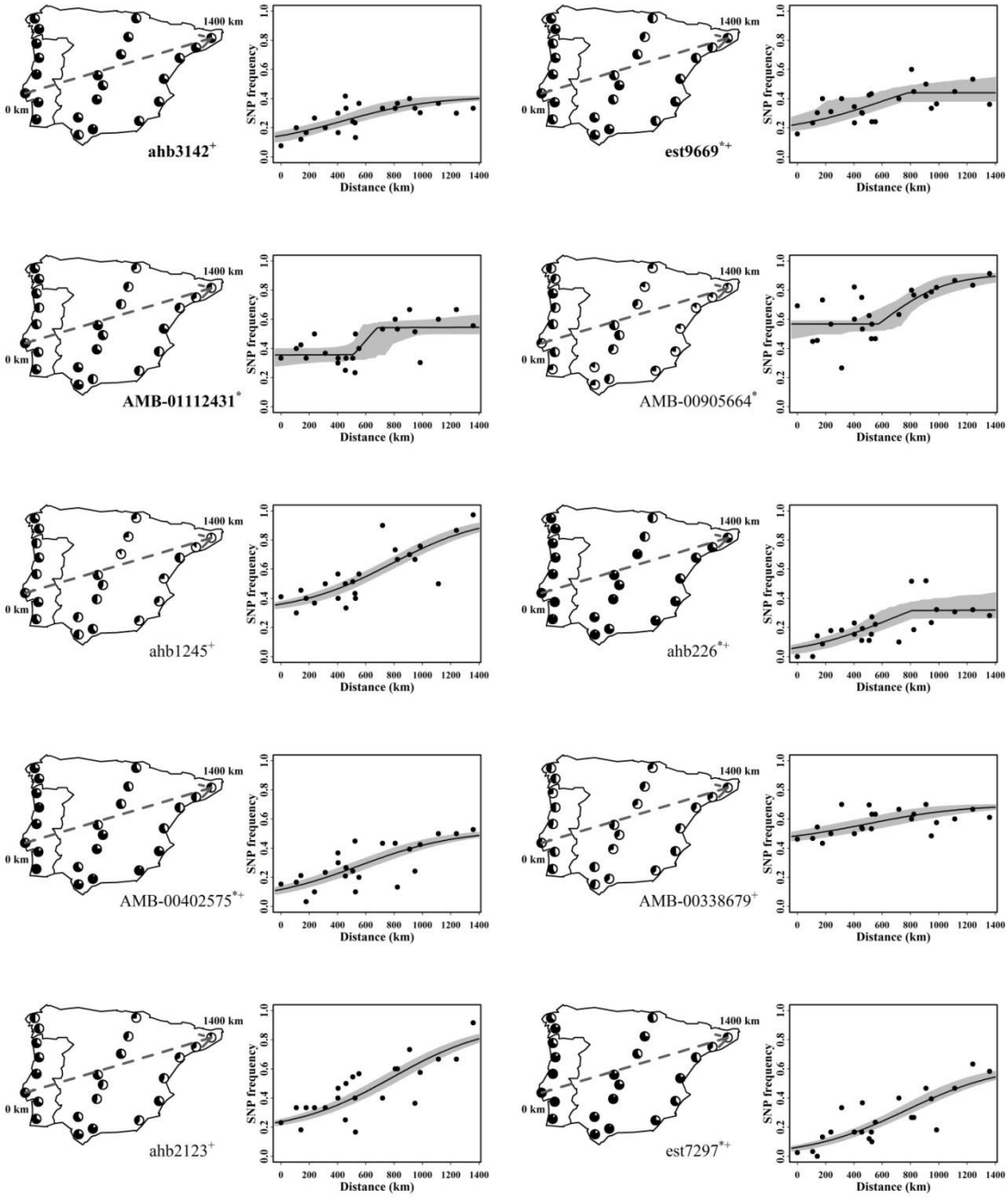


Figure Sup VI - 11. (Cont.)

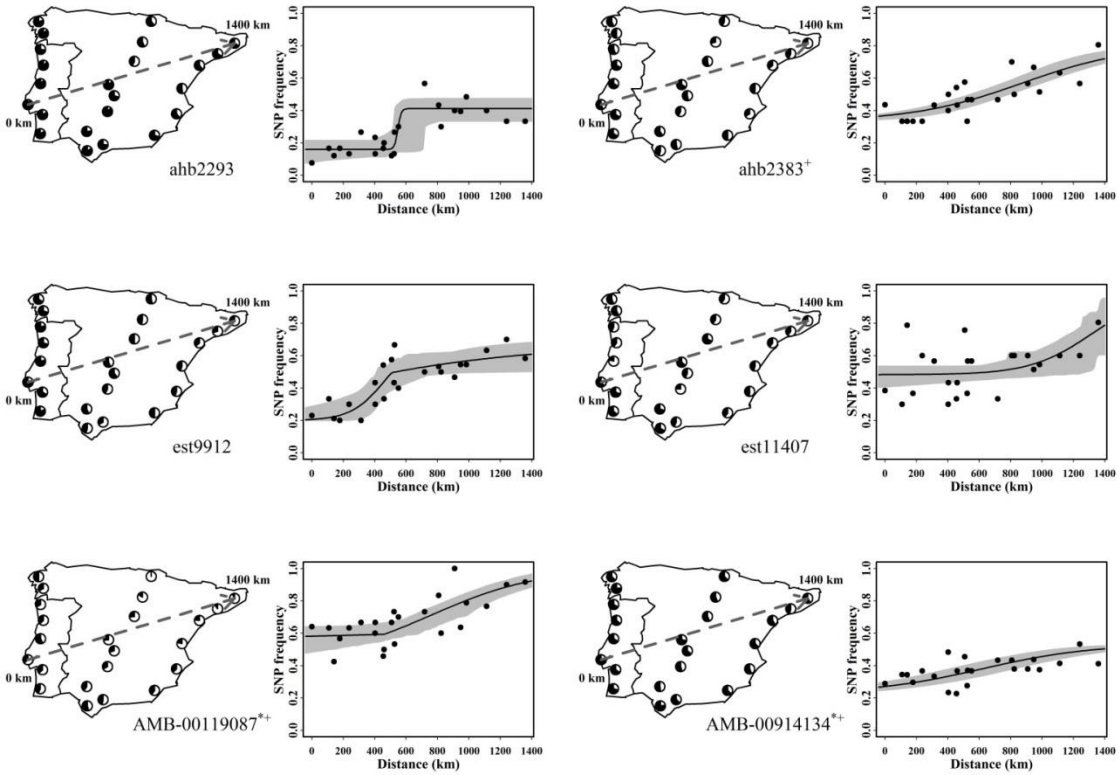


Figure Sup VI - 11. (Cont.)

Published Papers

Apidologie (2012) 43:663–673

DOI: 10.1007/s13592-012-0141-1

Original article

The Atlantic side of the Iberian Peninsula: a hot-spot of novel African honey bee maternal diversity

Maria Alice Pinto[1], Irene Muñoz[2], Julio Chávez-Galarza[1], Pilar De la Rúa[2]

[1]Mountain Research Centre (CIMO), Polytechnic Institute of Bragança, Campus de Sta. Apolónia, Apartado 1172, 5301-855 Bragança, Portugal
[2]Área de Biologia Animal, Dpto. de Zoología y Antropología Física, Facultad de Veterinaria, Universidad de Murcia, Campus de Espinardo, 30100 Murcia, Spain

Received 10 November 2011 – Revised 14 April 2012 – Accepted 2 May 2012

Abstract – The Iberian honey bee has been one of the most surveyed subspecies for genetic diversity. Yet, previous studies have missed an important component of Iberian honey bee variation harbored by populations inhabiting the Atlantic side of the Iberian Peninsula. Herein, we provide a fuller picture of the Iberian honey bee maternal diversity by revealing 16 novel haplotypes detected in honey bees from Portugal. Of the 16 haplotypes, all of African ancestry, 15 belong to the Atlantic sub-lineage A_{III} while only one fits the most common sub-lineage A_I. This level of new variation is remarkable as it represents a 59% increase in the wide-range African lineage and a 188% in sub-lineage A_{III}. Our findings further highlight the complexity of the Iberian honey bee diversity patterns and reinforce the importance of this southernmost European territory as a reservoir of *Apis mellifera* genetic diversity, a resource increasingly important in a rapidly changing and demanding world.

Iberian honey bee / genetic diversity / mtDNA / *Dra*I test / Portugal

1. INTRODUCTION

Understanding patterns and underlying processes of diversity of the western honey bee (*Apis mellifera* L.) has been a major goal of numerous genetic studies, which in a population declining scenario is becoming increasingly important because it can provide a stronger scientific basis for management and conservation decisions. Among the 29 recognized honey bee subspecies (Engel 1999; Sheppard and Meixner 2003), the Iberian honey bee has been one of the most intensively surveyed for diversity patterns. Indeed, studies of Iberian honey bees have been performed since the 1970s (Ruttner et al. 1978) using morphology (Cornuet and Fresnaye 1989; Arias et al. 2006; Miguel et al. 2010), allozymes (Smith and Glenn 1995; Arias et al. 2006), mitochondrial DNA (Smith et al. 1991; Garnery et al. 1995, 1998a; Franck et al. 1998; De la Rúa et al. 2001, 2004, 2005; Arias et al. 2006; Miguel et al. 2007; Cánovas et al. 2008), and microsatellites (Franck et al. 1998; Garnery et al. 1998b; De la Rúa et al. 2002, 2003; Miguel et al. 2007; Cánovas et al. 2011). Differential and complex patterns of diversity have emerged from these surveys, which have yet to be fully resolved.

The maternally inherited mitochondrial DNA (mtDNA) has been the marker of choice for

Electronic supplementary material The online version of this article (doi:10.1007/s13592-012-0141-1) contains supplementary material, which is available to authorized users.

Corresponding author: M.A. Pinto, apinto@ipb.pt
Manuscript editor: Marina Meixner

assessing Iberian honey bee variation, particularly the PCR–RFLP of the intergenic tRNA^{leu}-cox2 region, also known as the *Dra*I test (Garnery et al. 1993). Over 2,500 colonies, mostly sampled in the eastern half of the Iberian Peninsula, have been screened with the *DraI* test (Garnery et al. 1993, 1995, 1998a; Miguel et al. 2007; Cánovas et al. 2008). The data generated by this massive sampling confirmed early findings of coexistence of African (A) and western European (M) lineages, forming a south–north cline (Smith et al. 1991), and revealed unparalleled levels of haplotype diversity and complexity (Franck et al. 1998; Cánovas et al. 2008). Accordingly, it has been suggested that, as for other animal and plant species, the Iberian Peninsula served as a glacial refuge (Hewitt 1999, 2001; Arias et al. 2006; Gómez and Lunt 2007), and as a place of secondary contact between European and African evolutionary lineages (Smith et al. 1991; Garnery et al. 1995; Cánovas et al. 2008), and therefore it has been a stage for historical processes possibly involving recurring phenomena of local adaptation, contraction, fragmentation, expansion, and admixture.

In contrast with the populations inhabiting the eastern side of the Iberian Peninsula, Portuguese honey bees have been largely undersampled. Yet, the few samples collected in Portugal (Garnery et al. 1998a; Arias et al. 2006; Miguel et al. 2007) suggest that the Atlantic side of Iberia may harbor an important component of the Iberian honey bee maternal diversity. Therefore, a fuller understanding of diversity patterns of the Iberian honey bee requires further surveys of Ibero-Atlantic populations. As part of an ongoing genetic study of the honey bees occupying the Portuguese territory, we have detected 16 novel haplotypes with the *Dra*I test. This finding is non-trivial as this small corner of southern Europe is contributing with 59% of new variation to the wide-range African lineage. Herein, the novel haplotypes are fully described by the RFLP approach and by sequence data. Our findings suggest that the Atlantic side of the Iberian Peninsula harbors important genetic resources, especially in face of the escalating threats to the honey bee diversity.

2. MATERIALS AND METHODS

2.1. Samples and RFLP analysis

As part of an ongoing study of the Portuguese honey bee populations, over 950 stationary colonies were sampled, between 2008 and 2010, covering every district of continental Portugal and the archipelagos of Azores and Madeira (sample sizes and locations are provided in Online resource 1). Honey bee workers were collected from the inner part of the hives, placed in absolute ethanol, and then stored at −20°C until molecular analysis. The maternal ancestry of the 950 workers, each representing a single colony and a single apiary, was assessed using the *Dra*I test (Garnery et al. 1993), which consists on PCR amplification of the tRNA^{leu}-cox2 intergenic region, using the primers E2 and H2 (see PCR details in Garnery et al. 1993), followed by digestion with the restriction enzyme *Dra*I (see digestion and gel electrophoresis details in Cánovas et al. 2008). Of the 950 individuals scored using the complete set of restriction maps and restriction fragment sizes reported to date (De la Rúa et al. 1998, 2005; Franck et al. 2001; Collet et al. 2006), 43 exhibited a total of 16 unreported PCR–RFLP patterns. Location of the 43 colonies, from which the individuals were sampled, is shown in Figure 1. In this study, the 16 novel PCR–RFLP patterns were fully characterized by sequencing.

2.2. Sequencing and sequence analysis

The tRNA^{leu}-cox2 intergenic region analyzed herein contains a non-coding sequence which size depends on the forms of the P element and the number of repeats of the Q element. In the African lineage, the P element can display two different forms: P_0 (sub-lineage A_I and A_{II}) and P_1 (sub-lineage A_{III}). The P_0 differs from P_1 by a 15-bp deletion. The Q element can be repeated in tandem one to four times. A further distinction between sub-

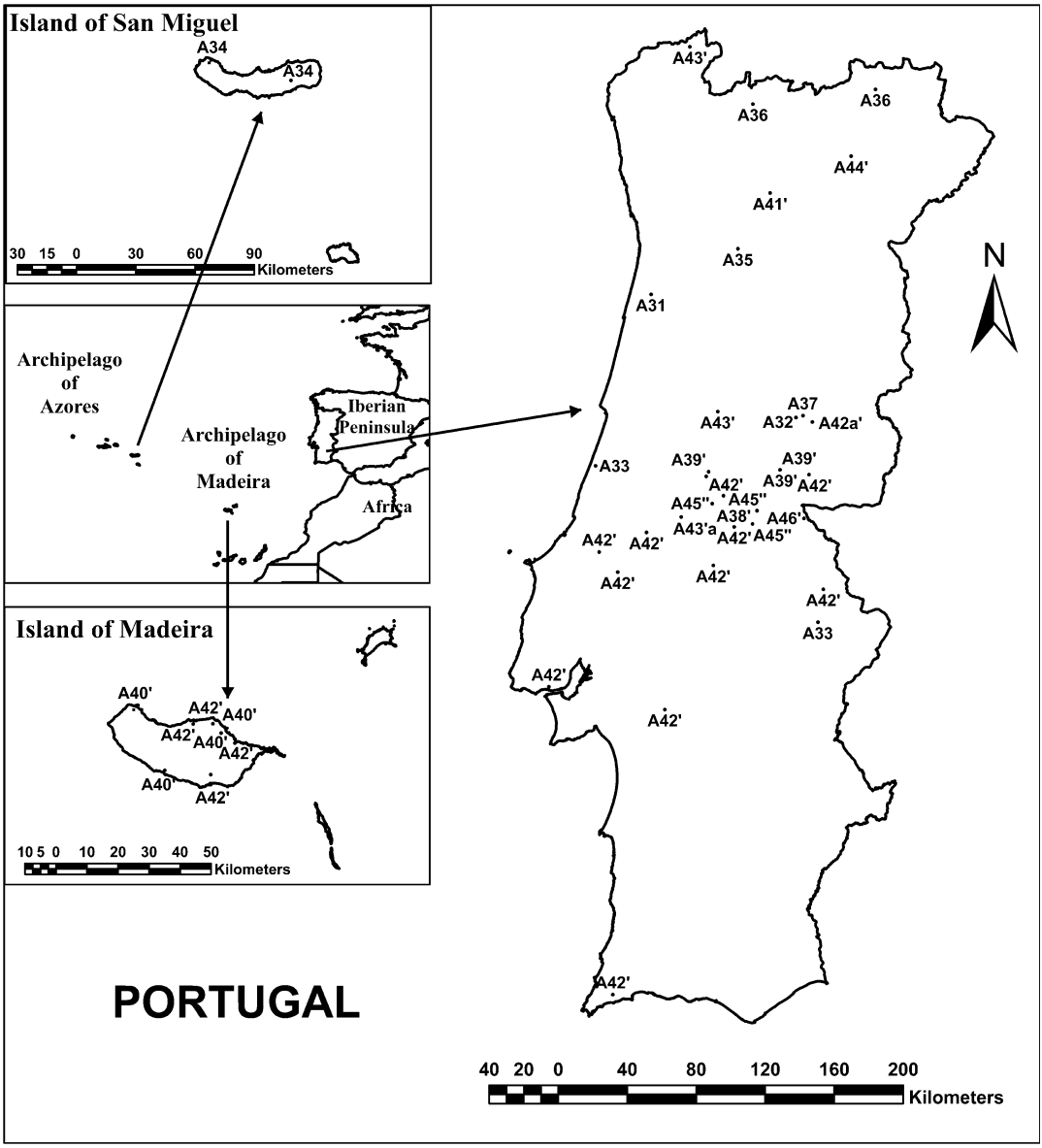


Figure 1. Location of the 43 colonies harboring the 16 novel PCR–RFLP haplotypes identified with the *Dra*I test.

lineages A_I and A_{II} is provided by the number of *Dra*I recognition sites in the region spanning $tRNA^{leu}$ and 5′ end of the first Q element, with two and one sites, respectively (Garnery et al. 1993; De la Rúa et al. 1998; Franck et al. 1998).

The 16 novel PCR–RFLP patterns were further examined by sequencing the $tRNA^{leu}$-cox2 intergenic region for 20 individuals. In addition, three individuals exhibiting the previously reported haplotypes A3 (sub-lineage A_I), A14, and A16 (sub-lineage A_{III}) were also sequenced for comparison. Following PCR amplification, PCR products were purified either with isopropanol and ammonium acetate or using a column-based purification kit

(Zymo Research®) and sent to Secugen S. L. (Madrid, Spain) or Macrogen (Seoul, Korea) for direct sequencing in both directions with primers E2 and H2. The sequences were checked for base calling using SeqMan® version 7.0.0 and then deposited in GenBank (http://www.ncbi.nlm.nih.gov).

The 23 sequences, plus the published sequences of haplotypes A29a (FJ890930.1; Szalanski and Magnus 2010) and A30 (EF033654.1; Collet et al. 2006), belonging to sub-lineage A_{III}, were aligned using MEGA version 5.03 (Tamura et al. 2011). Proximity among the haplotypes was established using the median-joining network algorithm (Bandelt et al. 1999) as implemented in the program Network version 4.6.0.0 (Fluxus Engineering, Clare, UK; http://www.fluxus-engineering.com), with epsilon set to zero and downweighting the four most variable sites (weight=0 for the most variable site; weight=5 for the remaining three most variable sites). Indels (e.g., P_1 element, third and fourth Q element, 15-deletion of A40′) were considered as a single mutational step, being therefore coded as a 1-bp gap. Variable characters within the first (Q1), second (Q2), and third Q (Q3) elements were included in the analysis.

3. RESULTS

The *Dra*I test performed on colonies surveyed across Portugal (continent and archipelagos of Azores and Madeira) identified 16 novel PCR–RFLP patterns (haplotypes) carried by 43 colonies (Figure 1). The restriction maps and length of restriction fragments (Figure 2) suggest that they all fit within the African evolutionary lineage (A). Following the nomenclature established earlier (Garnery et al. 1998a) and recently reviewed for lineage M (Rortais et al. 2011), the 16 haplotypes were numbered sequentially from A31 to A46 (accession numbers JQ746684–JQ746701). Haplotypes showing the same RFLP pattern but bearing three or four Q elements were further distinguished by addition of the symbols ′ and ″, respectively, after the haplotype number (Garnery et al. 1998a; Rortais et al. 2011).

Figure 2 shows that the 16 patterns are dramatically distinct from those previously reported for the African lineage. The number of restriction fragments varied between three (A31, A41′, A45″) and five (A32, A35, A37, A40′), with most haplotypes exhibiting a four-fragment pattern (A33, A34, A36, A38′, A39′, A42′, A43′, A44′, A46′). The shortest band (28 bp) was displayed by four haplotypes (A35, A36, A37, A39′) whereas the longest (1,064 bp) was unique to A45″.

Location of the 43 colonies, and corresponding distribution of the 16 haplotypes depicted in Figure 1, shows that haplotype A42′ was the most widespread and common (16 colonies) followed by A40′ (four colonies). Haplotypes A39′, A43′, and A45″ were carried by three colonies whereas A33, A34, and A36 by two colonies. The remaining eight haplotypes were singletons (A31, A32, A35, A37, A38′, A41′, A44′, and A46′). While most haplotypes were detected in the center of continental Portugal, A34 and A40′ were private to the islands of San Miguel (Azores) and Madeira, respectively. A42′ was the only haplotype detected in both mainland and island (Madeira) populations.

Sequence data (Figure 3) confirmed the novelty of the 16 haplotypes of African ancestry and allowed identification of two additional variants of haplotypes A42′ and A43′, which were distinguished by a lowercase letter (A42′a, A43′a), as suggested by Rortais et al. (2011). Among the 16 haplotypes, 15 contained the P_1 element whereas only one (A46′) exhibited the P_0 element. The length of the sequenced region ranged from 803 bp to 1,204 bp, depending on the number of Q elements and presence of indels. Fifteen haplotypes contained either two (A31, A32, A33, A34, A35, A36, A37) or three Q elements (A38′, A39′, A40′, A41′, A42′, A43′, A44′, A46′). Only one haplotype (A45″) displayed a sequence with four Q elements (Figure 2). As typical for this intergenic region, numerous indels of variable sizes were detected within the P and Q elements (Figure 3). In addition to the 15-bp deletion characteristic of the P_1 element (marked as "*d1*" in Figures 2 and 3), large indels were displayed by A31, A34, A40′, and A41′. The largest fragment (33 bp, marked as "c" in Figure 3) was inserted in the

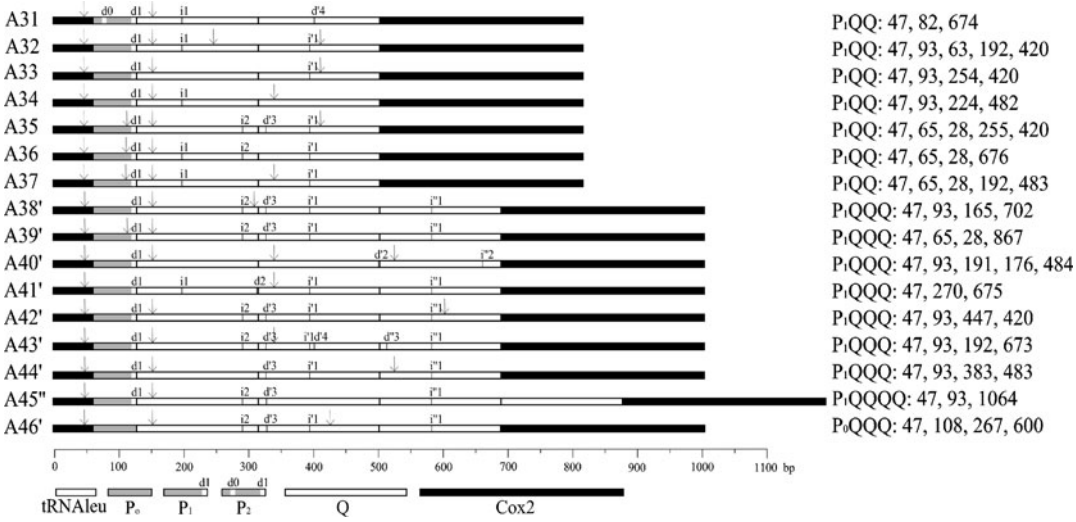


Figure 2. Restriction maps (*left*) and length of restriction fragments (*right*), deduced from *Dra*I restriction patterns and sequences of the tRNA^{leu}-cox2 intergenic region, of the 16 novel haplotypes found in Portugal. The *Dra*I recognition site is denoted with an *arrow*. Deletions and insertions are numbered preceded by the letters *d* and *i*, respectively. Deletion marked as *d1* is specific of the P_1 element which defines sub-lineage A_{III}, as in Franck et al. (2001).

Q1 element of haplotype A34. This 33-bp insertion produced the longest Q element (227 bp) ever reported. The 15-bp deletion motif of A40′ (marked as "l" in Figure 3) and A41′ (marked as "h" in Figure 3) was a perfect match of *d1*, the only difference being the position. Indeed, while *d1* was located at the 3′ end of the P element, deletions of A41′ and A40′ were at the 3′ end of Q1 and Q2, respectively. In addition to the variation originated from large indels, there were 10 short indels (1–2 bp) and 17 single-base substitutions, of which nine resulted in a gain/loss of the *Dra*I recognition site. The most striking mutation, displayed by seven haplotypes, is the "AG" inserted at the 3′ end of Q1 element (marked as "f" in Figure 3). Interestingly, except for the Q3 of haplotype A41′, no other Q element carried this mutation.

A median-joining network (Figure 4) based on 36 variable sites (18 coded indels and 18 substitutions) illustrates the relationships among the novel and previously described haplotypes of sub-lineages A_I (A3) and A_{III} (A14, A16, A29a, A30). Two distinct clusters, separated by the number of Q elements, are represented in the network. The more poorly resolved right-hand side cluster connects the haplotypes with two Q elements. One unresolved connection and four hypothetical haplotypes (unsampled or extinct) link this group to A35, the closest haplotype of the left-hand side cluster. Haplotypes A36 and A37 are the closest in the group with only one mutational step (a transition in A36 that led to loss of *Dra*I restriction site) separating them. The haplotype pairs A30/A37 and A32/A33 are separated by one hypothetical haplotype and two mutational steps (one is a transversion that led to a gain/loss of *Dra*I restriction site). The previously described A14 and A34, which are derived from the same hypothetical haplotype, are the most distant in the group. A minimum of eight mutations, including the duplication of the Q element, and two hypothetical nodes separate the closest haplotypes of both clusters (A35 and A43′a).

The better resolved left-hand side cluster connects the haplotypes with three and four Q elements. The 14 haplotypes are separated from each other by one to eight mutational steps and three hypothetical nodes. The previously de-

	76	77	78	79	80	81	82	83	84	85	86	102	111	112	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	158	197	198	199	200	201
	a											**1**	**2**	**3**	***d1***															4	c				
A3	A	T	A	A	A	A	T	A	A	A	A	T	T	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	A	-	-	-	-	-
A14	A	T	A	A	A	A	T	A	A	A	A	C	T	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-
A16	A	T	A	A	A	A	T	A	A	A	A	T	T	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-
A29a	A	T	A	A	A	A	T	A	A	A	A	T	T	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-
A30	A	T	A	A	A	A	T	A	A	A	A	T	T	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-
A31	-	-	-	-	-	-	-	-	-	-	-	T	T	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-
A32	A	T	A	A	A	A	T	A	A	A	A	T	T	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-
A33	A	T	A	A	A	A	T	A	A	A	A	T	T	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-
A34	A	T	A	A	A	A	T	A	A	A	A	C	A	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	T	T	T	A	T
A35	A	T	A	A	A	A	T	A	A	A	A	T	T	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-
A36	A	T	A	A	A	A	T	A	A	A	A	T	T	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-
A37	A	T	A	A	A	A	T	A	A	A	A	T	T	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-
A38'	A	T	A	A	A	A	T	A	A	A	A	T	T	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-
A39'	A	T	A	A	A	A	T	A	A	A	A	T	T	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-
A40'	A	T	A	A	A	A	T	A	A	A	A	T	T	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-
A41	A	T	A	A	A	A	T	A	A	A	A	T	T	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-
A42'	A	T	A	A	A	A	T	A	A	A	A	T	T	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-
A42'a	A	T	A	A	A	A	T	A	A	A	A	T	T	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-
A43'	A	T	A	A	A	A	T	A	A	A	A	T	T	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-
A43'a	A	T	A	A	A	A	T	A	A	A	A	T	T	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-
A44'	A	T	A	A	A	A	T	A	A	A	A	T	T	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-
A45''	A	T	A	A	A	A	T	A	A	A	A	T	T	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-
A46'	A	T	A	A	A	A	T	A	A	A	A	T	T	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	A	-	-	-	-	-

	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	252	275	303	304	326	327
																													d	5	6	e		7	f
A3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	A	-	-	G	A
A14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	C	A	T	G	-
A16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	C	-	-	G	A
A29a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	C	-	-	G	A
A30	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	T	C	-	-	G	-
A31	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	T	C	-	-	G	-
A32	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	A	C	-	-	G	-
A33	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	C	-	-	G	-
A34	T	T	T	T	A	T	A	T	T	T	A	A	T	A	A	A	A	T	A	A	A	T	A	A	T	T	A	A	-	T	C	-	-	G	-
A35	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	C	-	-	G	A
A36	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	T	C	-	-	G	-
A37	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	T	C	-	-	G	-
A38'	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	C	-	-	G	A
A39'	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	C	-	-	G	A
A40'	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	C	-	-	G	-
A41	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	T	C	-	-	A	-
A42'	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	C	-	-	G	A
A42'a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	C	-	-	G	A
A43'	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	C	-	-	G	A
A43'a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	C	-	-	G	A
A44'	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	C	-	-	G	A
A45''	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	C	-	-	G	A
A46'	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	A	-	-	G	A

Figure 3. Variable sites of tRNAleu-cox2 intergenic region of the 16 novel haplotypes, the two additional sequence variants (A42′a and A43′a), and the previously reported African haplotypes (A3, A14, A16, A29a, A30). Sequences of A29a (FJ890930.1) and A30 (EF033654.1) were obtained from GenBank. Sequences of A3, A14, and A16 were obtained from individuals collected in Portugal. The *numbers* on the *top line* represent the nucleotide position within the $tRNA^{leu}$-cox2 sequenced region. Position 1 marks the beginning of the $tRNA^{leu}$ gene, corresponding to position 3370 of the honey bee mitochondrial genome (Crozier and Crozier 1993). Substitution sites are numbered from 1 to 17. Indels (marked with a *dash*) are denoted by letters from *a* to *q*. The 15-bp deletion characteristic of the P_1 element is marked as *d1* (position 115). Variable sites between positions 158–358, 365–550, and 551–714 are within Q1, Q2, and Q3 elements, respectively. Position 745 identifies the beginning of Q4. Full sequences are available in GenBank under accession numbers JQ746684–JQ746701.

	328	329	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	365	387	406	429	435	448	457	471	536	537	538	539	540	541	542	543	544
		g	**h**												**8**			**i**	**9**	**10**	**j**	**k**	**11**	**12**	**13**	**l**								
A3	G	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	-	G	C	T	T	T	C	C	A	T	T	T	A	A	T	T	T
A14	-	-	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	C	A	T	-	T	T	C	C	A	T	T	T	A	A	T	T	T
A16	G	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	-	G	T	-	T	T	C	C	A	T	T	T	A	A	T	T	T
A29a	G	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	-	G	T	T	T	T	C	C	A	T	T	T	A	A	T	T	T
A30	-	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	C	G	T	T	T	T	C	C	A	T	T	T	A	A	T	T	T
A31	-	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	C	G	T	T	-	T	C	C	A	T	T	T	A	A	T	T	T
A32	-	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	C	G	T	T	T	A	C	C	A	T	T	T	A	A	T	T	T
A33	-	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	C	G	T	T	T	A	C	C	A	T	T	T	A	A	T	T	T
A34	-	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	C	A	T	-	T	T	C	C	A	T	T	T	A	A	T	T	T
A35	G	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	-	G	T	T	T	A	C	C	A	T	T	T	A	A	T	T	T
A36	G	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	C	G	T	T	T	T	C	C	A	T	T	T	A	A	T	T	T
A37	-	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	C	A	T	T	T	T	C	C	A	T	T	T	A	A	T	T	T
A38'	G	A	A	T	T	T	A	A	T	T	T	A	T	T	T	A	A	-	G	T	T	T	T	C	C	A	T	T	T	A	A	T	T	T
A39'	G	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	-	G	T	T	T	T	C	C	A	T	T	T	A	A	T	T	T
A40'	-	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	C	A	T	-	T	T	C	C	-	-	-	-	-	-	-	-	-
A41	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	A	T	T	T	T	C	C	A	T	T	T	A	A	T	T	T
A42'	G	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	-	G	T	T	T	T	C	T	A	T	T	T	A	A	T	T	T
A42'a	G	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	-	G	T	T	T	T	C	C	A	T	T	T	A	A	T	T	T
A43'	G	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	-	A	T	T	-	T	C	C	A	T	T	T	A	A	T	T	T
A43'a	G	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	-	A	T	T	T	T	C	C	A	T	T	T	A	A	T	T	T
A44'	-	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	-	G	T	T	T	T	C	C	A	T	T	T	A	A	T	T	T
A45''	G	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	-	G	T	-	T	T	C	C	A	T	T	T	A	A	T	T	T
A46'	G	A	A	T	T	T	A	A	T	T	T	A	T	T	A	A	A	-	G	C	T	T	T	T	C	A	T	T	T	A	A	T	T	T

	545	546	547	548	549	550	551	557	579	593	621	640	709	713	714	745
							m	n	14	15	o	16	17	p		q
A3	A	T	T	A	A	A	A	C	G	C	T	T	A	-	-	-
A14	A	T	T	A	A	A	-	-	-	-	-	-	-	-	-	-
A16	A	T	T	A	A	A	A	C	G	C	T	T	A	-	-	-
A29a	A	T	T	A	A	A	A	C	G	C	T	T	A	-	-	-
A30	A	T	T	A	A	A	-	-	-	-	-	-	-	-	-	-
A31	A	T	T	A	A	A	-	-	-	-	-	-	-	-	-	-
A32	A	T	T	A	A	A	-	-	-	-	-	-	-	-	-	-
A33	A	T	T	A	A	A	-	-	-	-	-	-	-	-	-	-
A34	A	T	T	A	A	A	-	-	-	-	-	-	-	-	-	-
A35	A	T	T	A	A	A	-	-	-	-	-	-	-	-	-	-
A36	A	T	T	A	A	A	-	-	-	-	-	-	-	-	-	-
A37	A	T	T	A	A	A	-	-	-	-	-	-	-	-	-	-
A38'	A	T	T	A	A	A	A	C	G	C	T	T	A	-	-	-
A39'	A	T	T	A	A	A	A	C	G	T	T	T	A	-	-	-
A40'	-	-	-	-	-	-	A	C	A	C	-	T	A	A	G	-
A41	A	T	T	A	A	A	A	C	G	C	T	T	A	-	-	-
A42'	A	T	T	A	A	A	A	C	G	C	T	A	A	-	-	-
A42'a	A	T	T	A	A	A	A	C	G	C	T	A	A	-	-	-
A43'	A	T	T	A	A	A	A	-	G	C	T	T	A	-	-	-
A43'a	A	T	T	A	A	A	A	-	G	C	T	T	G	-	-	-
A44'	A	T	T	A	A	A	A	C	A	C	T	T	A	-	-	-
A45''	A	T	T	A	A	A	A	C	G	C	T	T	A	-	-	A
A46'	A	T	T	A	A	A	A	C	G	C	T	T	A	-	-	-

Figure 3. (continued).

scribed A29a is central to this group, showing considerably more connections (eight) than any other haplotype in the network. This central haplotype is connected to the closest ones by a single mutation (1-bp deletion in Q2 for A16, and a substitution in Q3 for A42′a) and to the most distant (A41′) by eight mutational steps and one hypothetical node. The previously described A3 haplotype of sub-lineage A_I (as defined by Franck et al. 2001) and the novel A46′ are together in a branch that is separated from A29a by three and four mutations, respectively. These two sub-lineage A_I haplotypes differ from each other by one transition

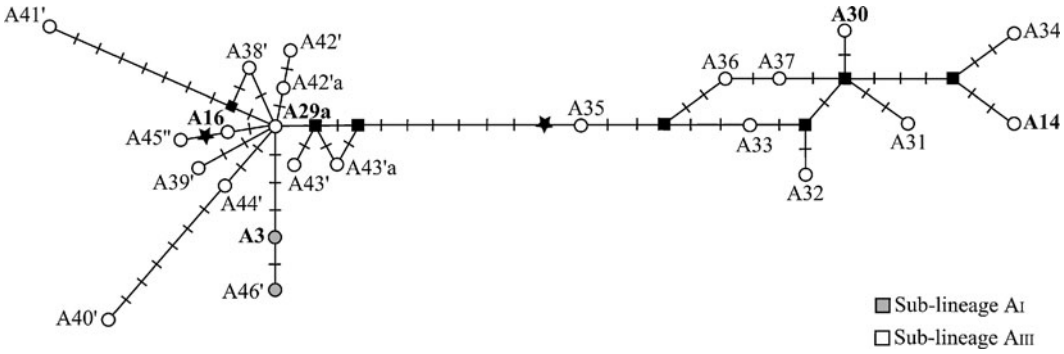


Figure 4. Median-joining network of the novel and previously described African haplotypes (marked in *bold*) that were identified from the tRNA^{leu}-cox2 region. Haplotypes are denoted as *circles*. Hypothetical (unsampled or extinct) haplotypes are symbolized as *filled squares*. The *cross lines* along branches represent mutational steps between nodes. The duplication of the Q element is represented by a *star*.

mutation (C/T), which accounted for an additional *Dra*I recognition site in Q2 of haplotype A46′ (substitution 12 in Figure 3).

4. DISCUSSION

This study further highlights the Atlantic side of the Iberian Peninsula as an important repository of Iberian honey bee maternal diversity. Analysis of Portuguese honey bee populations with the *Dra*I test revealed 16 novel haplotypes of African ancestry, which were confirmed by sequence data of the tRNA^{leu}-cox2 intergenic region. These haplotypes join the 27 African haplotypes previously described (De la Rúa et al. 1998, 2005; Franck et al. 2001; Collet et al. 2006) of which 17 have been found in the Iberian Peninsula (Garnery et al. 1998a; De la Rúa et al. 2004, 2005; Cánovas et al. 2008), representing an increase of 59% and 94%, respectively. Most haplotypes (15) contained the P_1 element, typical of sub-lineage A_{III}, whereas only one was assigned to sub-lineage A_I, as defined by Franck et al. (2001). The 15 haplotypes were added to the eight previously reported for sub-lineage A_{III} (De la Rúa et al. 1998; Franck et al. 2001; Collet et al. 2006), representing an increase of 188%.

The Iberian honey bee has been one of the most intensively surveyed subspecies in its natural range (Garnery et al. 1995, 1998a; Franck et al. 1998; Arias et al. 2006; Miguel et al. 2007; Cánovas et al. 2008). Therefore, detection of such a remarkable number of novel haplotypes, mostly of sub-lineage A_{III} ancestry, was unexpected and suggests that prior studies have missed an important diversity component hold by the western populations of Iberia. This study not only adds to the complexity of the Iberian honey bee diversity patterns (Franck et al. 1998; Cánovas et al. 2008) but also reinforces the Atlantic distribution proposed for sub-lineage A_{III} (De la Rúa et al. 1998, 2006; Franck et al. 2001), as most colonies were detected in the north of continental Portugal (Figure 1), which exhibits a more Atlantic climate contrasting with the more Mediterranean southern Portugal. Our findings further support an ancient natural colonization of the Iberian Peninsula by African swarms (De la Rúa et al. 2002, 2004; Cánovas et al. 2008) as the hypothesis of historical human-mediated multiple introductions (Franck et al. 1998; Garnery et al. 1998a) is untenable with such complex levels of diversity.

Among the 16 novel haplotypes, A42 is probably the oldest because of its higher frequency and wider geographical distribution (sole haplotype shared between mainland and island colonies). Alternatively, it could have been disseminated by human-assisted colony transportation, as occurred with A29. While the history of introductions of A29 (Collet et al.

2006; Prada et al. 2009) and its variant A29a (Szalanski and Magnus 2010) is unknown, we postulate that these haplotypes descend from colonies of Portuguese origin. The PCR–RFLP patterns of A29 (47/93/866) and A29a (47/93/867) are virtually indistinguishable from those of A16 (47/93/866). Additionally, sequence data shows that they are closely related (Figure 4 and sequences on GenBank), suggesting that A29 (Collet et al. 2006) and 29a (Szalanski and Magnus 2010) are merely sequence variants of A16. The PCR–RFLP haplotype A16 has only been reported in the Portuguese territory, including mainland (Garnery et al. 1998a; Miguel et al. 2007) and islands (De la Rúa et al. 2006). Therefore, it is possible that haplotypes A29 and A29a descend from Portuguese colonies introduced by settlers in South America in historical times and later expanded to North America (Szalanski and Magnus 2010) by Africanized honey bees. Alternatively, there were multiple independent, historical or recent, introductions in North and South America of putatively Portuguese-derived colonies.

Similarities in primary and secondary structures between the Q element and the 3′ end of COI gene (5′ end of Q), $tRNA^{leu}$ gene (middle part of Q), and the P sequence (3′ end of Q) led Cornuet et al. (1991) to propose an origin of the Q element by tandem replication. The first and the second Q of haplotypes A41′ and A40′, respectively, bear a 3′ end similar to P_1 whereas those of the other Q's are similar to P_0. Interestingly, the 3′ end of Q1 of some lineage M haplotypes (M34-HQ337456.1; M43Q-HQ260365.1) and Z haplotypes (Z1-HM236204.1; Z12-HM236212.1; Z13-HM236213.1; Z1Q-HM236205.1) also bear a deletion motif similar to the P_1 element. While this deletion has probably multiple independent origins, this finding deserves further investigation as it may shed some light in the evolution of this complex region.

This study further expands on the complexity of the Iberian honey bee patterns and reinforces the importance of this southernmost European territory as a reservoir of *A. mellifera* genetic diversity. In response to a rapidly changing world (e.g., new pests and parasites, land use change), which has had severe consequences in apiculture, there is a growing alert for protecting honey bee genetic resources across its natural range (Jensen et al. 2005; De la Rúa et al. 2009; Dietemann et al. 2009; Haddad et al. 2009) and an increasing number of conservation programs, specially to protect *A. m. mellifera* (reviewed by De la Rúa et al. 2009). Preservation of honey bee genetic variation is a pre-requisite for long-term adaptive change and avoidance of fitness decline, through inbreeding depression, and thereby a guarantee of a sustainable apiculture. The Iberian Peninsula has been a stage for evolutionary events that have shaped the evolutionary history of western European honey bee lineage. Therefore, this territory certainly deserves special attention in both small- and large-scale conservation programs.

ACKNOWLEDGMENTS

We are indebted to numerous technicians of the *Divisões de Intervenção Veterinária* (*Direcção Geral de Veterinária*) and of the Beekeeping Associations, who kindly provided the honey bee samples. Financial support was provided by *Fundação para a Ciência e Tecnologia* and COMPETE/QREN/EU (project PTDC/BIA-BEC/099640/2008), Programa Apícola Nacional (Medida 6A) to M. A. Pinto, and by *Fundación Séneca* (project 11961/PI/09) to P. De la Rúa. We also thank the two anonymous reviewers for their valuable comments.

La façade atlantique de la péninsule ibérienne : un « point chaud » pour une nouvelle diversité maternelle de type africain chez l'abeille

Abeille ibérienne / diversité génétique / ADN mitochondrial / test *Dra*I / Portugal

Die atlantische Seite der iberischen Halbinsel: ein Hotspot für bisher unbekannte Afrika-typische maternale Diversität bei Honigbienen

Iberische Honigbiene / genetische Diversität / mtDNA / DraI-Test / Portugal

REFERENCES

Arias, M.C., Rinderer, T.E., Sheppard, W.S. (2006) Further characterization of honey bees from the Iberian Peninsula by allozyme, morphometric and mtDNA haplotype analyses. J. Apic. Res. **45**, 188–196

Bandelt, H.J., Forster, P., Rohl, A. (1999) Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. **16**, 37–48

Cánovas, F., De la Rúa, P., Serrano, J., Galián, J. (2008) Geographical patterns of mitochondrial DNA variation in *Apis mellifera iberiensis* (Hymenoptera: Apidae). J. Zool. Syst. Evol. Res. **46**, 24–30

Cánovas, F., De la Rúa, P., Serrano, J., Galián, J. (2011) Microsatellite variability reveals beekeeping influences on Iberian honeybee populations. Apidologie. **42**, 235–251

Collet, T., Ferreira, K.M., Arias, M.C., Soares, A.E.E., Del Lama, M.A. (2006) Genetic structure of Africanized honeybee populations (*Apis mellifera* L.) from Brazil and Uruguay viewed through mitochondrial DNA COI–COII patterns. Heredity. **97**, 329–335

Cornuet, J.M., Fresnaye, J. (1989) Biometrical study of honey bee populations from Spain and Portugal. Apidologie. **20**, 93–101

Cornuet, J.M., Garnery, L., Solignac, M. (1991) Putative origin and function of the intergenic region between COI and COII of *Apis mellifera* L. mitochondrial DNA. Genetics. **128**, 393–403

Crozier, R.H., Crozier, Y.C. (1993) The mitochondrial genome of the honey bee *Apis mellifera:* complete sequence and genome organization. Genetics. **133**, 97–117

De la Rúa, P., Serrano, J., Galián, J. (1998) Mitochondrial DNA variability in the Canary islands honeybees (*Apis mellifera* L.). Mol. Ecol. **7**, 1543–1547

De la Rúa, P., Galián, J., Serrano, J., Moritz, R.F.A. (2001) Molecular characterization and population structure of the honeybees from the Balearic islands (Spain). Apidologie. **32**, 417–427

De la Rúa, P., Galián, J., Serrano, J., Moritz, R.F.A. (2002) Microsatellite analysis of non-migratory colonies of *Apis mellifera iberica* from south-eastern Spain. J. Zool. Syst. Evol. Research. **40**, 164–168

De la Rúa, P., Galián, J., Serrano, J., Moritz, R.F.A. (2003) Genetic structure of Balearic honeybee populations based on microsatellite polymorphism. Genet. Sel. Evol. **35**, 339–350

De la Rúa, P., Jiménez, Y., Galián, J., Serrano, J. (2004) Evaluation of the biodiversity of honey bee (*Apis mellifera*) populations from eastern Spain. J. Apic. Res. **43**, 162–166

De la Rúa, P., Hernandez-Garcia, R., Jiménez, Y., Galián, J., Serrano, J. (2005) Biodiversity of *Apis mellifera iberica* (Hymenoptera: Apidae) from northeastern Spain assessed by mitochondrial analysis. Insect Syst. Evol. **36**, 21–28

De la Rúa, P., Galián, J., Pedersen, B.V., Serrano, J. (2006) Molecular characterization and population structure of *Apis mellifera* from Madeira and the Azores. Apidologie. **37**, 699–708

De la Rúa, P., Jaffé, R., Dall'Olio, R., Muñoz, I., Serrano, J. (2009) Biodiversity, conservation and current threats to European honeybees. Apidologie. **40**, 263–284

Dietemann, V., Pirk, C.W.W., Crewe, R. (2009) Is there a need for conservation of honeybees in Africa?. Apidologie. **40**, 285–295

Engel, M.S. (1999) The taxonomy of recent and fossil honey bees (Hymenoptera: Apidae; Apis). J. Hym. Res. **8**, 165–196

Franck, P., Garnery, L., Solignac, M., Cornuet, J.M. (1998) The origin of west European subspecies of honeybees (*Apis mellifera*): new insights from microsatellite and mitochondrial data. Evolution. **52**, 1119–1134

Franck, P., Garnery, L., Loiseau, A., Oldroyd, B.P., Hepburn, H.R., Solignac, M., Cornuet, J.M. (2001) Genetic diversity of the honey bee in Africa: microsatellite and mitochondrial data. Heredity. **86**, 420–430

Garnery, L., Solignac, M., Celebrano, G., Cornuet, J.M. (1993) A simple test using restricted PCR-amplified mitochondrial DNA to study the genetic structure of *Apis mellifera* L. Experientia. **49**, 1016–1021

Garnery, L., Mosshine, E.H., Oldroyd, B.P., Cornuet, J.M. (1995) Mitochondrial DNA variation in Moroccan and Spanish honey bee populations. Mol. Ecol. **4**, 465–471

Garnery, L., Franck, P., Baudry, E., Vautrin, D., Cornuet, J.M., Solignac, M. (1998a) Genetic diversity of the west European honey bee (*Apis mellifera mellifera* and *A. m. iberica*). I. Mitochondrial DNA. Genet. Sel. Evol. **30**, S31-S47

Garnery, L., Franck, P., Baudry, E., Vautrin, D., Cornuet, J.M., Solignac, M. (1998b) Genetic diversity of the west European honey bee (*Apis mellifera mellifera* and *A. m. iberica*). II. Microsatellite loci. Genet. Sel. Evol. **30**, S49-S74

Gómez, A., Lunt, D.H. (2007). Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. Springer, Dordrecht

Haddad, N., Meixner, M.D., Fuchs, S., Migdadi, H., Garnery, L., Sheppard, W.S. (2009) Mitochondrial DNA support for genetic reserves of *Apis mellifera syriaca* in Jordan. J. Apic. Res. **48**, 19–22

Hewitt, G.M. (1999) Post-glacial re-colonization of European biota. Biol. J. Linn. Soc. **68**, 87–112

Hewitt, G.M. (2001) Speciation, hybrid zones and phylogeography—or seeing genes in space and time. Mol. Ecol. **10**, 537–549

Jensen, A.B., Palmer, K.A., Boomsma, J.J., Pedersen, B.V. (2005) Varying degrees of *Apis mellifera ligustica* introgression in protected populations of the black honeybee, *Apis mellifera mellifera*, in northwest Europe. Mol. Ecol. **14**, 93–106

Miguel, I., Iriondo, M., Garnery, L., Sheppard, W.S., Estonba, A. (2007) Gene flow within the M evolutionary lineage of *Apis mellifera*: role of the Pyrenees, isolation by distance and post-glacial re-colonization routes in the western Europe. Apidologie. **38**, 141–155

Miguel, I., Baylac, M., Iriondo, M., Manzano, C., Garnery, L., Estonba, A. (2010) Both geometric morphometric and microsatellite data consistently support the differentiation of the *Apis mellifera* M evolutionary branch. Apidologie. **42**, 150–161

Prada, Q.C.F., Duran, J.T., Salamanca, G.G., Del Lama, M.A. (2009) Population genetics of *Apis mellifera* L. (Hymenoptera: Apidae) from Columbia. J. Apic. Res. **48**, 3–10

Rortais, A., Arnold, G., Alburaki, M., Legout, H., Garnery, L. (2011) Review of the DraI COI–COII test for the conservation of the black honeybee (*Apis mellifera mellifera*). Conserv. Genet. Resour. **3**, 383–391

Ruttner, F., Tassencourt, L., Louveaux, J. (1978) Biometrical–statistical analysis of the geographic variability of *Apis mellifera* L. Apidologie. **9**, 363–381

Sheppard, W.S., Meixner, M.D. (2003) *Apis mellifera pomonella*, a new honey bee subspecies from Central Asia. Apidologie. **34**, 367–375

Smith, D.R., Glenn, T.C. (1995) Allozyme polymorphisms in spanish honeybees (*Apis mellifera iberica*). J. Hered. **86**, 12–16

Smith, D.R., Palopoli, M.F., Taylor, B.R., Garnery, L., Cornuet, J.M., Solignac, M., Brown, W.M. (1991) Geographical overlap of two mitochondrial genomes in Spanish honeybees (*Apis mellifera iberica*). J. Hered. **82**, 96–100

Szalanski, A.L., Magnus, R.M. (2010) Mitochondrial DNA characterization of Africanized honey bee (*Apis mellifera* L.) populations from the USA. J. Apic. Res. **49**, 177–185

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S. (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol. Biol. Evol. doi:10.1093/molbev/msr121

MOLECULAR ECOLOGY

Molecular Ecology (2013) **22**, 5890–5907 doi: 10.1111/mec.12537

Signatures of selection in the Iberian honey bee (*Apis mellifera iberiensis*) revealed by a genome scan analysis of single nucleotide polymorphisms

JULIO CHÁVEZ-GALARZA,* DORA HENRIQUES,* J. SPENCER JOHNSTON,† JOÃO C. AZEVEDO,* JOHN C. PATTON,‡ IRENE MUÑOZ,§ PILAR DE LA RÚA§ and M. ALICE PINTO*

*Mountain Research Centre (CIMO), Polytechnic Institute of Bragança, Campus de Sta. Apolónia, Apartado 1172, 5301-855 Bragança, Portugal, †Department of Entomology, Texas A&M University, College Station, TX 77843-2475, USA, ‡Department of Forestry and Natural Resources, Purdue University, 715 W State Street, West Lafayette, IN 4797-2061, USA, §Área de Biología Animal, Dpto. De Zoología y Antropología Física, Universidad de Murcia, Campus de Espinardo, 30100 Murcia, Spain

Abstract

Understanding the genetic mechanisms of adaptive population divergence is one of the most fundamental endeavours in evolutionary biology and is becoming increasingly important as it will allow predictions about how organisms will respond to global environmental crisis. This is particularly important for the honey bee, a species of unquestionable ecological and economical importance that has been exposed to increasing human-mediated selection pressures. Here, we conducted a single nucleotide polymorphism (SNP)-based genome scan in honey bees collected across an environmental gradient in Iberia and used four F_{ST}-based outlier tests to identify genomic regions exhibiting signatures of selection. Additionally, we analysed associations between genetic and environmental data for the identification of factors that might be correlated or act as selective pressures. With these approaches, 4.4% (17 of 383) of outlier loci were cross-validated by four F_{ST}-based methods, and 8.9% (34 of 383) were cross-validated by at least three methods. Of the 34 outliers, 15 were found to be strongly associated with one or more environmental variables. Further support for selection, provided by functional genomic information, was particularly compelling for SNP outliers mapped to different genes putatively involved in the same function such as vision, xenobiotic detoxification and innate immune response. This study enabled a more rigorous consideration of selection as the underlying cause of diversity patterns in Iberian honey bees, representing an important first step towards the identification of polymorphisms implicated in local adaptation and possibly in response to recent human-mediated environmental changes.

Keywords: *Apis mellifera iberiensis*, balancing selection, directional selection, F_{ST} outlier tests, genome scan, single nucleotide polymorphism

Received 20 April 2013; revision received 12 September 2013; accepted 19 September 2013

Introduction

Patterns of genetic diversity are shaped by genome-wide (e.g. random genetic drift, gene flow) and locus-specific (e.g. natural selection) processes. While the former details population demography and phylogenetic history, the latter helps identify genes that are important for fitness and adaptation (Luikart *et al.* 2003). Accordingly, disentangling genome-wide (neutral) from locus-specific (selected) variation is a fundamental goal in evolutionary biology because the outcome not only leads to more robust inferences of demographic history but also to identification of ecologically relevant genetic variation involved in local adaptation. In the context of rapid human-induced environmental change (e.g. habitat

Correspondence: M. Alice Pinto, Fax: +351 273 325 405;
E-mail: apinto@ipb.pt

fragmentation, climate change, introduction of novel parasites and diseases), such knowledge is becoming increasingly important for better managing and preserving genetic diversity (Allendorf & Luikart 2007).

Recent advances in computer technology and statistical genetic methods provide the tools for addressing that goal (reviewed by Luikart *et al.* 2003; Nielsen 2005; Storz 2005; Vasemägi & Primmer 2005; Helyar *et al.* 2011). Inspired by the original idea of Lewontin & Krakauer (1973) a family of statistical methods, known as the F_{ST}-based outlier tests, identify loci that exhibit frequencies significantly different than expected under neutrality for a given demographic model in genome scan data (Beaumont & Nichols 1996; Vitalis *et al.* 2001; Schlötterer 2002; Beaumont & Balding 2004; Foll & Gaggiotti 2008; Excoffier *et al.* 2009), thereby allowing separation of neutral from selected variation. The basic rationale is that genetic differentiation is higher for loci affected by directional selection (directional outliers) and lower for loci under balancing selection (balancing outliers) as compared to neutral variation. Those loci are usually called directional outliers and balancing outliers, respectively. Selection can be further validated as the cause of outlier behaviour by combining genetic and nongenetic data for the identification of environmental factors that might act as selective pressures (Foll & Gaggiotti 2006; Joost *et al.* 2007; Nielsen *et al.* 2009; Coop *et al.* 2010; Gomez-Uchida *et al.* 2011; Nunes *et al.* 2011; Prunier *et al.* 2011; Shimada *et al.* 2011; Tsumura *et al.* 2012) and by illuminating functional roles of selected loci (Luikart *et al.* 2003; Vasemägi & Primmer 2005; Prunier *et al.* 2011; Shimada *et al.* 2011; Lehtonen *et al.* 2012; Tsumura *et al.* 2012), a task that is facilitated for organisms with annotated genomes.

Owing to increasingly affordable and quicker genotyping of numerous loci scattered in the genome of numerous individuals, it is becoming increasingly popular to implement outlier tests at the genome and population-wide scales. This approach has revealed signatures of selection and, whenever supported by environmental and genomic information, has provided unprecedented insights into the ecological and molecular basis of adaptation in numerous taxa including plants (Prunier *et al.* 2011; Tsumura *et al.* 2012), fishes (Renaut *et al.* 2011; Shimada *et al.* 2011), birds (Lehtonen *et al.* 2012), pigs (Ai *et al.* 2013) and insects (De Jong *et al.* 2013), among others. Adding to the list, in the present study, a genome-wide scan using single nucleotide polymorphisms (SNPs) was conducted in the honey bee subspecies that is native to the Iberian Peninsula: the Iberian honey bee (*Apis mellifera iberiensis*).

Apis mellifera iberiensis is one of 30 currently recognized subspecies of honey bees, which occur naturally in the Middle East, Africa and Europe (Ruttner 1988; Engel 1999; Sheppard & Meixner 2003; Meixner *et al.* 2011). Contrary to the African continent, where most honey bee subspecies are predominantly wild (Dietemann *et al.* 2009), in Europe, following arrival of the invasive mite *Varroa destructor* in early 1980s, wild colonies have virtually disappeared (Moritz *et al.* 2007; Jaffe *et al.* 2010). Native European honey bee subspecies (including the Iberian honey bee) are now mostly confined to apiaries.

The Iberian honey bee has been the subject of numerous population genetic surveys and thus represents one of the best studied, and yet controversial, among all subspecies. Maternal and biparental genetic markers have revealed highly complex and incongruent patterns of variation (Smith *et al.* 1991; Garnery *et al.* 1995, 1998a,b; Smith & Glenn 1995; Franck *et al.* 1998; Cánovas *et al.* 2008, 2011), which have led to competing hypotheses for the origin of the Iberian honey bee.

Early phylogeographical studies of morphology (Ruttner *et al.* 1978; Cornuet & Fresnaye 1989) and the allozyme malate dehydrogenase (Nielsen *et al.* 1994; Smith & Glenn 1995; Arias *et al.* 2006) revealed the existence of a smooth gradient extending from North Africa to France with Iberian honey bees showing intermediate phenotypes. This pattern raised the hypothesis of primary intergradation and an African origin for this subspecies (Ruttner *et al.* 1978). However, mitochondrial polymorphisms showed the co-occurrence of highly divergent African-derived (lineage A) and western European-derived (lineage-M) haplotypes forming not a smooth but a steep south–north cline in the Iberian Peninsula (Garnery *et al.* 1995; Franck *et al.* 1998; Arias *et al.* 2006; Miguel *et al.* 2007; Cánovas *et al.* 2008; Pinto *et al.* 2013), a pattern that was more consistent with a secondary contact scenario (Smith *et al.* 1991). Adding to the complexity, microsatellites (Franck *et al.* 1998; Garnery *et al.* 1998b; Cánovas *et al.* 2011; Miguel *et al.* 2011) and recent geometric morphometric data (Miguel *et al.* 2011) exhibited virtually no differentiation and no traces of African genes in Iberian honey bee populations and revealed a sharp break between Iberian and northern African populations, thereby supporting neither hypothesis. The difference between maternal and bi-parental variation led Franck *et al.* (1998) to reject the secondary contact hypothesis and propose historical human-assisted introductions of African colonies with selection the best explanation for the reported morphological and allozymic clines and the diffusion and maintenance of African haplotypes in the south-western half of the Iberian Peninsula. A recent study using SNPs suggests, however, that while selection may have shaped the genome of lineage-M honey bees, which includes *A. m. iberiensis* and *Apis mellifera mellifera*, the

process occurred during ancient expansions from Africa into Western Europe, resurrecting the primary intergradation hypothesis (Zayed & Whitfield 2008).

Selection has repeatedly been invoked to explain cytonuclear latitudinal patterns (Franck *et al.* 1998; Garnery *et al.* 1998a,b), yet no previous effort has attempted to evaluate its relative importance in structuring Iberian honey bee populations. In this study, populations sampled across three north–south Iberian transects were subjected to SNP-based genome scans to evaluate the importance of selection in shaping diversity patterns of Iberian honey bees. The ecological and molecular bases of the genomic regions exhibiting signatures of selection were further supported and investigated by spatial analysis, through nonrandom associations between locus-specific variation and environmental variables, and by functional annotations of genes marked by outlier SNPs. The approach followed here revealed unprecedented insights into Iberian honey bee diversity patterns, which might be helpful for future management and conservation of honey bees, an increasingly important endeavour given the current worldwide concern regarding honey bee health.

Methods

Sampling

A total of 711 honey bee haploid males, representing 23 sites (Fig. 1) and 237 apiaries, were collected in 2010 across three north–south transects in the Iberian Peninsula. The sites were selected to represent both the natural distribution of *Apis mellifera iberiensis* and a wide variety of climates ranging from the semi-arid in south-eastern to oceanic in north-western Iberia. One transect extended along the Atlantic coast (AT, 8 sites), one through the centre (CT, 9 sites) and another along the Mediterranean coast (MT, 6 sites). The number of apiaries sampled per site varied between 8 and 13, with most sites having 10. Accordingly, sample size per site varied between 24 (8 apiaries per site × 3 hives per apiary) and 39, with most sites having 30 individuals. In each apiary, samples were taken from the inner part of three different hives and placed into absolute ethanol. Samples were stored at −20 °C until molecular analysis. Global positioning system (GPS) coordinates were recorded in the field for each apiary.

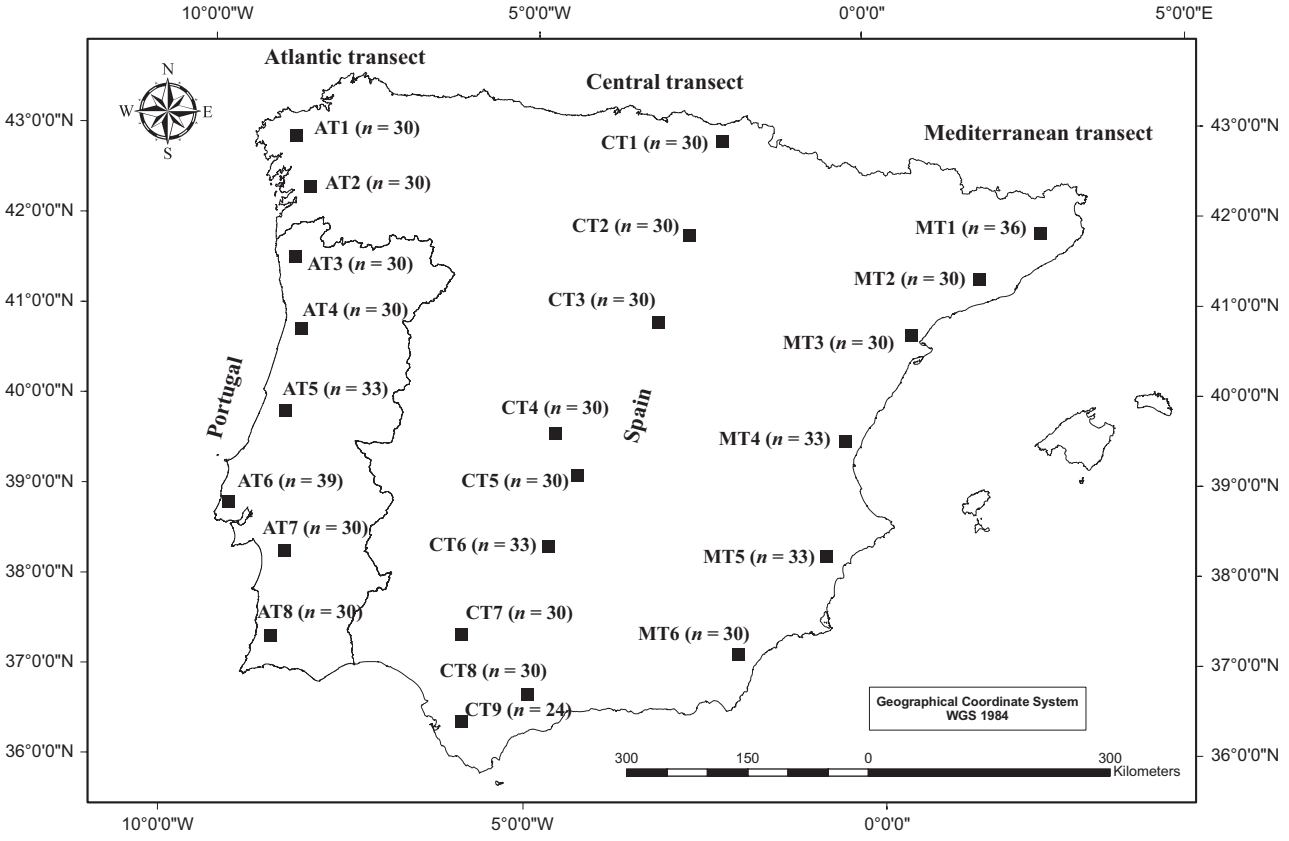


Fig. 1 Map of the Iberian Peninsula showing the centroids of the sampling sites, sample size per site, and site codes. The number of apiaries sampled per site varied between 8 and 13, with most sites having 10.

DNA extraction and SNP genotyping

Total DNA was extracted using a phenol/chloroform isoamyl alcohol (25:24:1) protocol (Sambrook *et al.* 1989) from the thorax of 711 individuals, each representing a single colony. A total of 1536 SNP loci were genotyped for those individuals using Illumina's BeadArray Technology and the Illumina GoldenGate® Assay with a custom Oligo Pool Assay (Illumina, San Diego, CA, USA) following manufacturer's protocols.

The Oligo Pool consisted of the 768 most informative SNPs used previously to study honey bee population structure and evolution (Whitfield *et al.* 2006) combined with a newly developed set of 768 SNPs. Both sets of 768 SNPs were drawn from over 1.1 million SNPs defined by single base differences between (i) the reference genome of *A. mellifera* (Assembly 3.0; sequenced from the North American DH4 strain, which was primarily *A. m. ligustica*) and genome sequence traces of Africanized honey bees (largely *A. m. scutellata* admixed with the genomes of both western and eastern European honey bees) and (ii) observed polymorphisms in ESTs. In the first case, SNPs were named 'ahb' and 'AMB', whereas in the second, they were named 'est' (Table 1 and Table S1, Supporting information). The new 768 SNPs were selected with the goal of obtaining markers that were evenly spaced across the honey bee genome (D. Weaver, personal communication).

Genotype calling was performed using Illumina's GenomeStudio® Data Analysis software. For each sample, intensity clusters generated automatically by the software were manually verified and edited when necessary. SNPs with poorly separated clusters or low signals (110) were excluded from the data set. For the remaining 1426 SNPs, most honey bee samples (695 of 711) exhibited a call rate between 95% and 100%. The rest of the samples (16) had a call rate lower than 95% but above 90%.

Environmental data

Publicly available environmental data were obtained for the location of each apiary. Altitude was estimated using 30 arc-second (~1 km) spatial resolution data from the WorldClim database (http://www.worldclim.org). Climatic data were extracted from two data sets. The first data set (spatial resolution of 0.5°, representing ~50 km), which covered the period 1901–2009, was obtained from the Climatic Research Unit (www.cru.uea.ac.uk), Norwich, UK, and consisted of precipitation (Prec), minimum temperature (T_{min}), mean temperature (T_{mean}), maximum temperature (T_{max}) and cloud cover (Cld). The second data set (spatial resolution of 1°, representing ~100 km), which covered the period 1983–2005 (http://eosweb.larc.nasa.gov), was downloaded from OPENEI (http://en.openei.org), and consisted of relative humidity (Rh) and insolation (Ins), which is the amount of radiation reaching the Earth's surface per day (insolation on horizontal surface in $kWh/m^2/day$). All climatic data were integrated into a geographic information system (ArcGIS 9.3 from ESRI) to extract yearly, seasonal and monthly data. Land use/land cover data for the Iberian Peninsula was extracted from the CORINE Land Cover 2006 vector data from the European Environment Agency (http://www.eea.europa.eu). Land cover was described for each apiary by calculating the percentage of level 3 land cover classes (Heymann *et al.* 1994) within a 3 km radius circular area (28.3 km^2). To remove redundant environmental variables, that is, variables that were correlated at $|r| > 0.8$ (Manel *et al.* 2010), a principal component analysis was performed using the *ade4* package (Thioulouse *et al.* 1997). Using this procedure, we kept 80 environmental variables (Table S2, Supporting information) for further analysis, from an initial data set of 123.

Detection of outlier loci by F_{ST}-based methods

Outliers were detected using four multiple-population F_{ST}-based methods. These methods assume varying demographic models to identify loci under selection as outliers in the extreme tails of theoretical null distributions of F_{ST}. The purpose of employing conceptually different approaches was to identify potential false positives. The first method was FDIST2 (Beaumont & Nichols 1996) as implemented in LOSITAN (Antão *et al.* 2008). This coalescence-based method uses an island model to identify as outlier loci those that present unusually low or high F_{ST} values compared with neutral expectations. Approximation to the mean neutral F_{ST} in the data set was accomplished by choosing the neutral mean F_{ST} option (99% confidence interval) and running 1 000 000 simulations in LOSITAN. The second method, implemented in Arlequin 3.5.1.3, is a modification of FDIST2 that overcomes possible false positives in the presence of strong population structure by using a hierarchical island model (Excoffier *et al.* 2009). The hierarchical outlier analysis was performed by pooling the sites into an Atlantic group (all AT sites), an M-lineage group (CT1–CT3, MT1–MT3) and an A-lineage group (CT4–CT9, MT4–MT6), based on maternal patterns (Fig. S1, Supporting information) and analysis of molecular variance (Table 2, Fig. S2, Supporting information). The parameters of the run included the presence of 50 groups of 100 demes with 100 000 iterations simulated. The third method, implemented in BAYESFST, uses Markov Chain Monte Carlo simulations to assess the significance of a locus-specific

Table 1 Outlier SNP loci as detected by four and three F_{ST}-based methods (BAYESFST, Arlequin and LOSITAN with *P*-values ≤ 0.05; BAYESCAN with posterior $P \geq 0.95$) and that are associated with environmental variables identified by the spatial analysis method (matSAM; CI ≥99%)

SNP code	Linkage group	Position	Gene product	Putative function	SNP location	matSAM
Loci detected by BAYESFST, Arlequin, LOSITAN, and BAYESCAN						
Directional						
ahb1245	1	25131116	Gst-mic2—microsomal glutathione-S-transferase 2	Metabolism	Intron	Long, Prec, T_{min}, Ins
est5302	6	3199025	UDP-glucosyltransferase (UDP-glucosyltransferase 35b, Ugt35b)*	Metabolism	Exon	Lat, Prec, T_{mean}, T_{max}, Cld
est5112	6	6341643	Vha16—Vacuolar H+ ATP synthase 16 kDa proteolipid subunit (Vacuolar H^+ATPase subunit 16-1, Vha16-1)*	Transport	3′-UTR	
ahb8266	6	13484985	Teneurin 3—like isoform 1 (Tenascin major, Ten-m)*	Structural	Intron	Lat, Prec, T_{mean}, T_{max}, Cld, Ins
ahb10181†	9	7162173	Hypothetical protein LOC726750/Hydrocephalus- inducing protein-like	Unknown/ Structural	4110/1297	
ahb2123	10	2249429	Choline transporter-like protein 1-like	Transport	3′-UTR	Long, Lat, Prec, Ins
est7297	10	3151832	15-hydroxyprostaglandin dehydrogenase [NAD+]-like (Photoreceptor dehydrogenase, Pdh)*	Metabolism	Intron	Long, Lat, Prec, Ins
ahb2105	10	6427742	Hypothetical protein LOC100577401	Unknown	Exon	Long, Prec, Rh
AMB-00644533	11	8036193	bs-Serum response factor homologue (blistered, bs)*	Regulation	Intron	Long, Prec, Ins
est9898	13	1328979	GTP-binding protein CG1354 isoform 1	Signalling	Exon	Lat, Prec, T_{mean}, T_{max}, Ins
est10016	13	9554011	Cytochrome P450-CYP6AS7 (Cyp6a14)*	Metabolism	Exon	Lat, Prec, T_{mean}, T_{max}, Cld, Ins
est11018	15	5299405	NimC2-nimrod C2 (nimrod C2, nimC2)*	Immunity	5′-UTR	
Balancing						
ahb142	1	21550552	Sema 1-Semaphorin 1A (Sema-1a)*	Structural	Intron	
ahb6903	4	9847583	Dscam-Down syndrome cell-adhesion molecule (Down syndrome cell-adhesion molecule, Dscam)*	Immunity	Intron	
AMB-00963630†	12	1214828	Hypothetical protein LOC100576488/Collagen alpha-2(IX) chain-like	Unknown/ Structural	18439/25703	
AMB-00708602	15	8856798	Cubilin-like	Signalling	Intron	
AMB-00914134‡	0	78294	Protein lin10-like (X11Lβ)*	Regulation	6219	
Loci detected by BAYESFST, Arlequin, and LOSITAN						
Directional						
AMB-00905664	1	15234278	Hypothetical protein LOC100578389	Unknown	Intron	
ahb1232†	1	25169032	Notum pectinacetylesterase homologue (Notum)*/NMDA kainate 2 sensitive receptor	Signalling/ Signalling	21596/10461	
ahb226§	1	29654719				
est2423	2	2983426	Retinol dehydrogenase 11-like	Metabolism	Exon	Ins
AMB-00190928	2	7469407	Ubc-E2H-Ubiquitin-conjugating enzyme E2 H (UbcE2H)*	Signalling	Intron	Lat, Prec, Ins
AMB-00402575	3	5547433	Protein outspread-like (outspread, osp)*	Regulation	Intron	
est5553	7	4067968	Aldh-Aldehyde dehydrogenase isoform 1 (Aldehyde dehydrogenase, Aldh)*	Metabolism	5′-UTR	Long
est6087	8	5157656	Rfabg-Retinoid and fatty acid-binding glycoprotein isoform1 (Retinoid- and fatty acid-binding glycoprotein, Rfabg)*	Transport	Exon	

Table 1 *Continued*

SNP code	Linkage group	Position	Gene product	Putative function	SNP location	matSAM
ahb9731	8	8266264	Hypothetical protein LOC411273	Unknown	Exon	Lat, Prec, Ins
ahb10154	9	7170480	Hydrocephalus- inducing protein-like	Structural	Exon	Lat, Prec, T_{mean}, Ins
AMB-00119087[§]	0	30896				Long, Prec
Balancing						
ahb1129	1	18156863	Hypothetical protein LOC413562	Regulation	Intron	
est5796	7	5788241	Hypothetical LOC100578906	Unknown	5′-UTR	
est6265	8	6383127	PHD finger and CXXC domain-containing protein CG17446-like isoform1 (Cfp1)*	Regulation	3′-UTR	
AMB-00338679	8	11725179	Midasin-like	Structural	Intron	
AMB-00310216[‡]	10	5583063	5 HT2 beta-Serotonin receptor	Signalling	22473	
ahb4188	14	9203355	Blop-blue-sensitive opsin (Rhodopsin, Rh5)*	Signalling	Exon	

SNP loci exhibiting the strongest signal (*P*-values ≤ 0.005 and posterior $P \geq 0.99$) are marked in bold. Genomic information for the SNP loci listed was obtained from NCBI (http://www.ncbi.nlm.nih.gov/), BEEBASE (http://hymenopteragenome.org/beebase), and FLYBASE (www.flybase.org).

Lat, latitude; Long, longitude; Prec, precipitation; T_{min}, minimum temperature; T_{mean}, mean temperature; T_{max}, maximum temperature; Ins, insolation; Cld, cloud cover; Rh, relative humidity.

*Names and/or symbols within parentheses correspond to orthologous genes of *Drosophila melanogaster* as in FLYBASE.

[†]SNP located between two genes (or putative genes). Function of both genes and physical distances (bp) to the 3′ or 5′ ends are indicated.

[‡]SNP located close to a gene. Physical distance (bp) to the 3′ or 5′ end of the gene is indicated in the column 'SNP location'.

[§]SNP located far away from genes (>167.8 kb).

parameter that indicates selection in a model of F_{ST} (Beaumont & Balding 2004). This Bayesian method was performed using 3 000 000 iterations. The fourth method, implemented in BAYESCAN 2.01, uses a Bayesian approach and a reversible-jump Markov Chain Monte Carlo method to estimate the posterior probability that a given locus is under selection (Foll & Gaggiotti 2008). It tests two alternative models: one that includes the effect of selection and another that excludes it. The BAYESCAN analysis was conducted using 20 pilot runs of 100 000 iterations, 1 500 000 iterations (sample size of 75 000 and thinning interval of 20) and an additional burn-in of 500 000 iterations.

Introgression and ascertainment bias

Because introgression may mimic selection, prior to the F_{ST}-based tests the 711 Iberian individuals were assessed for introgression using Structure 2.3.3 (Pritchard *et al.* 2000) and by implementing a principal components analysis with ADEGENET 1.3–7 (Jombart 2008). Over 1075 polymorphic SNPs and a reference collection consisting of the two eastern European beekeepers-favourite honey bee subspecies, *Apis mellifera ligustica* (17 individuals) and *Apis mellifera carnica* (19 individuals) and the northern African subspecies *Apis mellifera intermissa* (31 individuals) were used. These analyses revealed virtually no signs of contemporaneous or historical introgression from the two eastern European or the northern African subspecies (Figs S3 and S4, Supporting information for the parameters settings and results), suggesting that introgression should not be a confounding factor in the F_{ST}-based outlier detection.

Retaining uninformative monomorphic loci in the data set tends to increase dramatically the rate of false positives when searching for selection using F_{ST}-based methods (Nielsen *et al.* 2009; Gomez-Uchida *et al.* 2011). Accordingly, the Iberian data set was screened for monomorphic SNPs, as defined by a cut-off criterion of >0.98 for the most common allele. This filtering process produced a final data set of 383 polymorphic loci (288 genomic and 95 EST-derived) for the Iberian honey bee, which was used in all subsequent analyses performed in this study. The high proportion of monomorphic SNPs in Iberian honey bees might be explained by ascertainment bias. While the SNP panel was relatively diverse it did not include *A. m. iberiensis*. Accordingly, when SNPs were genotyped on Iberian honey bees, an ascertainment bias was introduced. However, ascertainment bias is expected to affect every genotyped Iberian individual equally and thus not systematically bias any particular individual or population, and thus evidence for selection is not caused by a SNP discovery artefact.

Table 2 Variance components (%) of the three-level AMOVAs obtained with the 309 neutral loci and the nine strongest directional outliers for different grouping combinations (all *P-values* < 0.0001)

		Within sites		Among sites		Among groups	
Grouping criteria	No. of groups	Neutral	Outlier	Neutral	Outlier	Neutral	Outlier
By mtDNA lineage*	2	95.75	80.14	3.35	11.56	0.89	8.29
By transect†	3	96.01	81.24	3.24	8.05	0.75	10.71
By AT & mtDNA lineage‡	3	95.92	80.70	3.05	6.59	1.03	12.70
By AT & mtDNA lineage§	3	95.95	81.32	3.09	8.17	0.96	10.51
By transect & mtDNA lineage¶	6	96.05	82.19	2.89	5.53	1.06	12.28

Grouping by transects and mtDNA lineage were the main criteria (see Fig. S2, Supporting information for a graphical visualization of the five groupings).
*Group 1: all M-lineage sites across transects; Group 2: all A-lineage sites across transects.
†Group 1: all AT sites; Group 2: all CT sites; Group 3: all MT sites.
‡Group 1: all AT sites; Group 2: M-lineage sites of CT and MT; Group 3: A-lineage sites of CT and MT.
§Group 1: only A-lineage AT sites; Group 2: M-lineage sites of AT, CT and MT; Group 3: A-lineage sites of CT and MT.
¶Group 1: M-lineage site of AT; Group 2: A-lineage sites of AT; Group 3: M-lineage sites of CT; Group 4: A-lineage sites of CT; Group 5: M-lineage sites of MT; Group 6: A-lineage sites of MT.

Genomic information

Each outlier SNP's 100 bp flanking sequence was mapped to the Honey Bee Assembly 4.5 using BLAST in BEEBASE (hymenopteragenome.org/beebase) and NCBI (www.ncbi.nlm.nih.gov). Genomic position was ascertained using the Map Viewer tool available in NCBI. SNPs were classified as belonging to exons, introns, 3′ or 5′ untranslated region (UTR), or intergenic regions. Genes marked by SNPs were identified using the Official Gene Set 3.2 (BEEBASE) and Entrez Gene (NCBI). As functional annotation of the honey bee genome is incomplete, putative Gene Ontology classifications were ascribed to as many genes as possible, based on homology to *Drosophila melanogaster*, using best-BLASTP hit and *e*-value cut-off 0.01 in FLYBASE (www.flybase.org) complemented by NCBI annotation.

Differentiation of neutral and directional outlier loci

Differentiation of putatively neutral and the strongest directional outliers was investigated by analysis of molecular variance (AMOVA). Two- and three-level AMOVAs were performed using Arlequin (Excoffier & Lischer 2010) with 10 000 permutations to assess whether levels of differentiation were significantly >0. Two-level AMOVAs (a single group) were conducted to test the prediction of a higher variance attributed to sites for directional outliers as compared to putatively neutral loci. Three-level AMOVAs (multiple groups) were conducted to assess (i) whether putatively neutral and outlier variation were similarly structured and, if there were any structure, (ii) whether neutral and/or outlier structure were oriented predominantly east–west or north–south. To that end, each neutral and directional outlier data set was partitioned by transect (longitudinal grouping) and mtDNA lineage (latitudinal grouping) into different grouping combinations (Fig. S2, Supporting information for details) in an effort to search for structures that generated the highest variance fraction attributed to groups.

Associations between environmental variables and outlier loci

A spatial analysis was performed to identify associations between allelic frequencies and environmental variables by using the software matSAM (Joost *et al.* 2008). This approach may provide insightful clues about selective forces acting upon outlier loci. Multiple univariate logistic regression models were computed by matSAM at the individual level (representing a single hive). The software matSAM assesses the significance of the coefficients calculated by the logistic regression function by implementing likelihood ratio (G) and Wald statistical tests. A model is considered significant only if the null hypothesis is rejected by both tests, after Bonferroni correction (Joost *et al.* 2007). The 383 SNP data set was tested against 80 environmental variables described previously (see Table S2, Supporting information). The significance threshold level was set to 1.632E−7, corresponding to a 99% confidence interval following Bonferroni correction. In addition to the matSAM analysis performed at the apiary level, associations between land cover and the strongest directional outlier loci were further examined for the 23 sites through linear regressions of allelic frequencies and percentage of level 3 land cover classes within a circular

area of 1963.5 km^2 (25 km radius) around each site's centroid. In this analysis, different combinations of agriculture land cover classes were tested, and only the best models were selected.

Results

Detection of outlier loci by F_{ST}-based methods

The genome scan approach implemented in this study identified, by at least one of the four F_{ST}-based methods, a total of 69 outlier loci (of 383; 18.0%) at a 95% confidence level (Table 1 and Table S1, Supporting information). The detection rate of outliers varied among the four methods, with the highest number of loci obtained by LOSITAN (57 of 383; 14.9%), and the lowest by BAYESCAN (17 of 383; 4.4%). The approaches implemented by Arlequin and BAYESFST detected 49 (12.8%) and 41 loci (10.7%), respectively. All loci detected by BAYESCAN were also identified by the other three methods, whereas 15 loci were exclusive to LOSITAN, six to Arlequin, and four to BAYESFST. As predicted from other studies (Wilding *et al.* 2001; Bonin *et al.* 2006; Shimada *et al.* 2011; Wang *et al.* 2012), the number of outlier loci exhibiting a signal of directional selection (50) was higher than that of balancing selection (19; Table 1 and Table S1, Supporting information) and varied according to the F_{ST}-based method employed. Specifically, the number of outlier loci under directional and balancing selection was as follows, respectively: LOSITAN 43/14, Arlequin 33/16, BAYESFST 27/14 and BAYESCAN 12/5 (Table 1, Table S1 and Fig. S5, Supporting information).

The 69 outliers were dispersed throughout the 16 honey bee linkage groups (LG) with three loci still unplaced in a chromosome (Fig. 2). LG 1, 2 and 6 harboured the highest number of outlier loci with 7, 10 and 7, respectively. The lowest physical distance (linkage) between outlier loci was observed for pairs ahb1245/ahb1232 (LG1) and ahb10181/ahb10154 (LG9) with 37.9 and 8.3 kb, respectively. The remaining 65 loci were further apart, with physical distances varying between 65.9 and 26 304 kb for pairs est9938/est9912 (LG13) and est1222/ahb226 (LG1), respectively. Therefore, except for the two pairs above, given the honey bee's exceptionally high recombination rate of 23.2 cM/Mb (Beye *et al.* 2006) most outlier loci might be considered unlinked. Indeed, in an earlier study using 1136 SNPs (many employed herein) across 14 honey bee subspecies and Africanized honey bees, Whitfield *et al.* (2006) observed a rapid decay of linkage disequilibrium over a distance of 5–10 kb.

Among the 69 outlier loci, 34 (23 directional and 11 balancing) were detected by at least three F_{ST}-based methods (Table 1), whereas 35 were detected by two or fewer F_{ST}-based methods (Table S1, Supporting information) and might be false positives. In contrast, 17 loci are strong candidates for selection as they were simultaneously detected by the four methods (Table 1). Among these 17, 10 loci (nine directional and one balancing) exhibited the strongest signal (P-value ≤ 0.005 and posterior $P \geq 0.99$) being therefore the best candidates (marked in bold in Table 1). The nine best directional outliers exhibited large differences in allele frequencies among sites, contrasting with the sole candidate (AMB-00963630) for balancing selection with nearly even

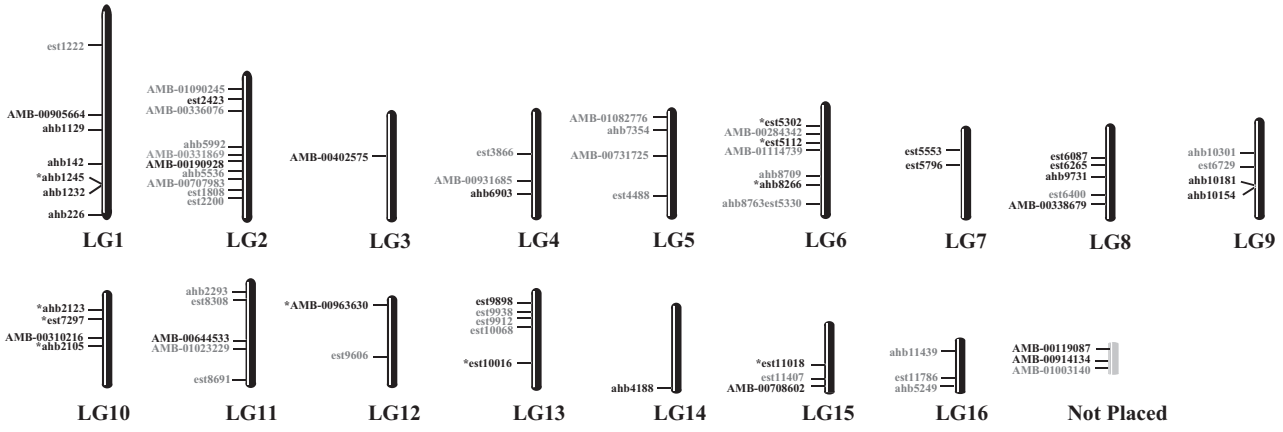


Fig. 2 Physical map of the 16 honey bee linkage groups (LG1 to LG16) showing the genomic positions of the 69 outlier SNP loci detected by at least one F_{ST}-based method and that are associated with at least one environmental variable. The 34 outliers detected by at least three F_{ST}-based methods are marked in bold (Table 1), whereas those detected by two or fewer F_{ST}-based methods (Table S1, Supporting information) are marked in grey. The 10 strongest candidates for selection (Table 1) are marked with an asterisk. Three SNPs have no position assigned yet. The map was depicted from the honey bee genome sequence available at http://www.ncbi.nlm.nih.gov/projects/mapview using the Map Viewer tool.

frequencies across the geographical range (Fig. S1, Supporting information). In spite of the weaker signal, 17 additional loci can still be considered good candidates as they were detected by three F_{ST}-based methods (Table 1).

Genomic information shows that of the 34 outliers, 32 loci were located in or near genes (Table 1) that code for proteins (Table S3, Supporting information for Accession nos.) involved in a diverse array of putative functions including signalling, structural, metabolism, regulation, transport and immunity. Among the 27 outliers that mapped to genes, 12 were located in introns, 6 in untranslated regions (3′ or 5′ UTR) and 9 in exons, although none was predicted to induce amino acid changes.

Differentiation of neutral and directional outlier loci

To assess whether putatively neutral and directional outlier variation were similarly structured, two- and three-level hierarchical AMOVAs were performed for different loci combinations (neutral vs. the strongest outliers). For the two-level AMOVA, the percentage of the total variance among sites was substantially higher for outliers (15.97%; $P < 0.0001$) than for neutral loci (3.77%; $P < 0.0001$). As observed for the two-level AMOVA, when multiple groups were considered (three-level AMOVA), the variance component attributed to groups was higher for outlier loci than for neutral loci, regardless of the groupings tested (Table 2).

For neutral loci, the variance due to groups was low (0.75–1.06%) and always smaller than the variance among sites (2.89–3.35%). The best variance partitioning was obtained for six groups (1.06), although slightly better (1.03) than that obtained for three groups formed by the Atlantic sites, the M-lineage sites of central and Mediterranean transects and the A-lineage sites of central and Mediterranean transects (Table 2, Fig. S2, Supporting information). Separating sites in the central and Mediterranean transects, using the same criteria applied to the Atlantic sites, generated poorer variance partitioning (data not shown). Overall, these results indicate a very weak neutral substructure across the Iberian Peninsula, although the Atlantic populations seem to be slightly more differentiated.

In contrast to the result with neutral loci, when the strongest directional outliers were analysed, the variance component attributed to groups was elevated (10.51–12.70%) and was higher than the variance among sites (5.53–8.17%), as long as the Atlantic transect was kept as a separate group. Grouping the Atlantic with central and/or Mediterranean sites always inflated the variance attributed to sites and lowered the among groups component (Table 2 and data not shown). In summary, these results indicate that while neutral structure is very weak, directional outlier structure is pronounced and stronger longitudinally than latitudinally.

Associations between environmental variables and outlier loci

In addition to the F_{ST}-based outlier approaches, genomic regions exhibiting a signal of selection were further confirmed by the spatial analysis implemented by matSAM at the individual level. Significant associations (≥99% confidence level with Bonferroni correction), with at least one environmental variable, were detected for 33 (8.62%) out of the 383 screened loci (Table S2, Supporting information). Of the 33 loci, five were exclusive to matSAM and 28 had previously been identified by at least one F_{ST}-based method as directional outliers (Table 1 and Table S1, Supporting information).

The environmental variables that were more frequently associated with SNPs were precipitation and longitude, with 21 and 17 such associations, respectively (Table 1, Tables S1 and S2, Supporting information). The variables altitude and land cover were on the opposite side of the spectrum with no significant (after Bonferroni correction) associations detected. Most loci (20) were associated with a single or two variables. A few loci (four) were associated with more than four variables. The greatest number and strongest associations were found for locus ahb8266, which was correlated with latitude, precipitation, minimum temperature, mean temperature, maximum temperature, cloud cover and insolation (Table 1).

Among the 34 loci that were detected by at least three and four F_{ST}-based methods, 15 and 9 loci, respectively, showed significant (≥ 99%) associations with at least one environmental variable (Table 1). Overall, only nine loci were simultaneously detected by the five methods, with confidence levels above 99.99%. These loci were the strongest candidates for directional selection and were mainly associated with latitude, longitude, precipitation and insolation.

Regression analyses of allele frequencies and land cover implemented at the site level revealed significant associations for six of the nine strongest directional outliers (Table 3). The strongest correlations were detected for loci est10016, est11018, and est5112. Loci ahb1245, ahb2123 and est7297 were not associated with any of the land cover classes tested. The land classes that, individually or combined, produced the best models were all agricultural and included 'nonirrigated arable land', 'permanently irrigated land', 'fruit trees and berry plantations', 'olive groves', 'annual crops associated with permanent crops', 'complex cultivation patterns' and 'agro-forestry areas'. None of the forest classes tested

Table 3 Best linear regression models between the nine strongest directional outliers (see Table 1) and land cover at the site scale

SNP code	Land cover classes*	r	r^2	P-value
ahb1245	None	—	—	—
ahb2123	None	—	—	—
est7297	None	—	—	—
ahb2105	222	0.4330	0.1875	0.03902
est11018	222	0.6415	0.4116	0.00097
est5112	211	0.5846	0.3417	0.00339
est5302	211 + 212 + 222 + 223 + 241	0.5391	0.2907	0.00794
ahb8266	211 + 212 + 222 + 223 + 241 + 242	0.5567	0.3099	0.00580
est10016	211 + 212 + 222 + 223 + 241 + 242 + 244	0.6591	0.4344	0.00062

*Corine land cover 2006 codes: (211) non-irrigated arable land, (212) permanently irrigated land, (222) fruit trees and berry plantations, (223) olive groves, (241) annual crops associated with permanent crops, (242) complex cultivation patterns, and (244) agro-forestry areas. For each site, land cover classes were considered individually and combined within a 20 km radius circle. Regression analyses tested for associations between allele frequencies and the land cover proportion for the 23 sites.

was found to be associated with allele frequencies (data not shown).

Discussion

Candidate SNP loci for selection

Determining whether an outlier is a marker of a selected locus or a false positive is a major concern when searching for adaptive molecular variation. A common strategy has been to seek for confirmatory evidence from multiple outlier approaches that generate the expected neutral distribution of F_{ST} estimates under different demographic scenarios (Luikart *et al.* 2003; Storz 2005; Vasemägi & Primmer 2005). In this study, of the 69 detected outliers, only 17 (all identified by BAYESCAN) were cross-validated by the four conceptually different methods, whereas 34 and 44 (Table 1 and Table S1, Supporting information) were simultaneously detected by at least three and two methods, respectively. While discrepant results among outlier approaches have been repeatedly reported (Bonin *et al.* 2006; Shikano *et al.* 2010; Nunes *et al.* 2011; Shimada *et al.* 2011; De Jong *et al.* 2013), comparative simulations have shown that BAYESCAN outperforms under a wide range of scenarios, exhibiting the lowest rate of false positives (Pérez-Figueroa *et al.* 2010; Narum & Hess 2011). Accordingly, these 17 outliers are the best candidates for selection. At the same time, the proportion of outlier loci detected by at least four (4.4%), three (8.9%) or two (11.5%) methods is within the range reported for other taxa (extensively reviewed by Nosil *et al.* 2009; Shikano *et al.* 2010; Nunes *et al.* 2011; Shimada *et al.* 2011; Buckley *et al.* 2012; among others), including honey bees (Zayed & Whitfield 2008), and all are potentially important markers, including 10 loci identified by only two methods that map to functional variation, which deserve further attention to avoid the risk of losing interesting candidates.

A common feature of outlier approaches is their higher power in detecting directional selection compared with balancing selection, a disparity that is particularly pronounced when levels of genetic differentiation are low (Beaumont & Balding 2004; Foll & Gaggiotti 2008). While this would explain a higher proportion of directional than balancing outliers detected herein and in most studies (Nielsen *et al.* 2009; Cooke *et al.* 2012; Limborg *et al.* 2012), the fact that five loci were identified by every single method in a scenario of low differentiation (F_{ST} values ranged between 0.008 and 0.093 for the 383 loci; data not shown) across the Iberian honey bee range indicates a strong signal of balancing selection, especially for locus AMB-00963630, which was detected with a confidence level above 99%.

Biological relevance of candidate genes and association with possible selection pressures

Support for selection comes from functional annotations of genes carrying outlier SNPs that relate directly to colony fitness, and their association with possible selection pressures identified by the spatial analysis (Table 1). Support is particularly compelling for outlier loci mapped to different genes putatively involved in the same function, which is the case of balancing and directional outliers that are mapped to genes encoding proteins that are related to vision, xenobiotic detoxification and immune response.

Outlier SNPs mapped to genes related to vision. Five outlier SNPs mark genes associated with vision. The strong directional outlier ahb8266 is mapped to an intron of a gene that encodes a transmembrane glycoprotein named teneurin 3-like isoform 1 (GB12816). Its putative

orthologue in *Drosophila melanogaster*, tenascin major, has been shown to accomplish an important role during neural development and in the process of vision (Kinel-Tahan *et al.* 2007). Balancing outlier ahb142 is also located in an intronic region of a transmembrane protein named Semaphorin 1A (GB11468), which is required for the synapse formation and axon guidance (Godenschwege *et al.* 2002). This protein has been shown to participate in regulating the photoreceptor axon guidance in the visual system of *D. melanogaster* establishing an appropriate topographic termination pattern in the optic lobe (Cafferty *et al.* 2006). Three additional outliers, two directional (est7297, est2423) and one balancing (ahb4188), are putatively related to vision through their participation in the visual cycle, which has been shown to be a major pathway contributing to the maintenance of rhodopsin levels in *D. melanogaster* (Wang *et al.* 2010). The strong outlier est7297 is located in the intronic region of a gene that encodes a 15-hydroxyprostaglandin dehydrogenase [NAD+]-like protein (GB11685) belonging to NADB-Rossman superfamily, based on sequence comparison. Its putative orthologue in *D. melanogaster* is a photoreceptor dehydrogenase that participates in regenerating the chromophore for the production of rhodopsin (Wang *et al.* 2010). Outlier est2423 is located in the exon of a gene that encodes a retinol dehydrogenase 11-like protein (GB11195), which might be implicated in the production of the chromophore by catalysing the *cis*-retinol to *cis*-retinal (Kim *et al.* 2005; Belyaeva *et al.* 2009; Wang *et al.* 2010). Finally, balancing outlier ahb4188 is mapped to the exon of a gene encoding for a blop (blue sensitive) opsin (GB13493), which accomplishes a receptor function of the chromophore and forms a rhodopsin responsible for absorbing the 440 nm (blue spectra) wavelength (Townson *et al.* 1998; Earl & Britt 2006). Fixation of an alternative allele could confer better protein activity or structural conformation (Camps *et al.* 2007) for the production and regeneration of the chromophore, while the maintenance of polymorphism could provide an adaptive advantage favouring the absorbance of different wavelengths within blue spectra, as recently reported for titi monkeys (Bunce *et al.* 2011).

The three directional outliers (ahb8266, est7297, and est2423) are associated with insolation, among other environmental variables, suggesting that these loci mark genomic regions that are involved in local vision-associated adaptation. Vision is a key component of foraging behaviour (Winston 1987). Flight departures from the colony for foraging have been both positively and negatively correlated with solar radiation intensity (Burrill & Dietz 1981). Once outside, foragers rely on their highly developed trichromatic visual system and use colour discrimination for finding food sources and homing to the hive (reviewed by Menzel & Müller 1996). It is therefore possible that selection favours alleles that enable a more efficient light perception at contrasting climates such as the Mediterranean with long periods of sunny skies in southern Iberia, and the Atlantic with long periods of rain and cloudy skies, particularly in north-western Iberia.

Outlier SNPs mapped to genes related to xenobiotic detoxification. Three directional SNPs mark genes involved with detoxification of xenobiotics. Outlier est10016 is located in the exon of a gene that encodes CYP6AS7 protein (GB18052), which belongs to the cytochrome P450s monooxygenases superfamily. This protein superfamily plays a major role in the protection against xenobiotics and has been implicated in tolerance to plant toxins (Scott & Wen 2001; Mao *et al.* 2009) and evolved resistance to pesticides in many insects (Feyereisen 1999; Ffrench-Constant *et al.* 2004; Li *et al.* 2007), including tolerance of pyrethroid insecticides in honey bees (Pilling *et al.* 1995; Johnson *et al.* 2006; Mao *et al.* 2011). Outlier ahb1245 is located in an intron of a putative gene that codes for a microsomal glutathione-S-transferase 2 (GB10566), which belongs to another major superfamily associated with detoxification, the glutathione-S-transferases (GSTs). This gene has been implicated in the detoxification of metabolites formed by cytochrome P450 enzymes (Yu 2002; Claudianos *et al.* 2006). Finally, outlier est5302 maps to the exon of a gene (GB10566) that codes for a protein belonging to the UDP-glycosyltransferases (UGT) superfamily whose probable function is related to olfaction and detoxification mechanisms, based on its putative *D. melanogaster* UDP-glucosyltransferase 35b orthologue (Bull & Whitten 1972; Wang *et al.* 1999). In addition to the xenobiotic metabolism, these genes might also play a role in the defence against pathogens. A study in *D. melanogaster* suggested that genes encoding GSTs and cytochromes P450 likely participate in the detoxification of reactive oxygen species produced during microbial killing to protect the gut epithelium (Buchon *et al.* 2009). A similar mechanism might be involved in honey bees, as suggested by an elevated GST activity in the honey bee gut infected by one of its most important pathogens, the microsporidium *Nosema ceranae* (Dussaubat *et al.* 2012).

It makes perfect sense that three outlier SNPs map to loci involved with detoxification. During their lifetime, honey bees are exposed to a wide range of natural (e.g. phytochemicals present in nectar, pollen and propolis) and synthetic (agricultural pesticides) xenobiotics. While honey bees have evolved detoxification mechanisms to metabolize natural phytochemicals, they are known for being unusually sensitive to a range of insecticides (Hardstone & Scott 2010). Agricultural

insecticides used for crop protection are increasingly recognized as major drivers of recent worldwide honey bee losses. In 2012 alone, more than 100 papers and reports were published on the impact of insecticides, particularly neonicotinoids, on bees (mentioned by Osborne 2012). In addition to acute poisoning accidents (e.g. the case in Germany described by vanEngelsdorp & Meixner 2010), neonicotinoids seem to have a pervasive impact on honey bees through short- and long-term chronical exposure to sub-lethal doses that can result in reduction of disease resistance and breeding success as well as behavioural disturbances including problems in flying and navigation, impaired memory and learning, and reduced foraging ability (extensively reviewed by Blacquière *et al.* 2012). Accordingly, honey bees are potentially under strong selection pressure when they forage in intensive farming areas. In this study, we observed significant correlations between agricultural land cover (a surrogate of agrochemicals use) and outlier SNPs est10016 and est5302 (Table 3). Given increasing evidence of sub-lethal effects of pesticides on colony survival, this finding deserves further investigation.

Outlier SNPs mapped to genes related to innate immune response. Two strong outlier SNPs, which showed contrasting selection signatures, mark genes implicated in immune response. While est11018 exhibits a signal of directional selection and ahb6903 is seemingly under balancing selection, both modes of selection have been shown to be associated with immune response in fruit fly (Jiggins & Hurst 2003; Sackton *et al.* 2007), cod (Gomez-Uchida *et al.* 2011) and humans (Hollox & Armour 2008). Directional outlier est11018 maps to the 5′UTR region of a gene that encodes a putative Nimrod C2 protein (GB13979). Nimrod C2 belongs to a diverse class of transmembrane proteins that have been shown to function as phagocytosis receptors and/or microbial binding factors, suggesting an important role in the cellular immunity and elimination of apoptotic cells (Kurucz *et al.* 2007; Sackton *et al.* 2007; Somogyi *et al.* 2008). Balancing outlier ahb6903 is located in the intron of a gene encoding a Dscam-Down syndrome cell-adhesion molecule (GB15141). *DSCAM*, an orthologous gene in *D. melanogaster*, encodes for a receptor that has long been implicated in neuronal development, although more recently it has also been shown to participate in the phagocytosis of infectious non-self (Watson *et al.* 2005). As expected, the two outliers were not associated with any of the environmental variables tested in this study. Despite potential direct influence of climate on both host and pathogen, selection pressure on genomic regions related to immune response is imposed by parasites and pathogens, which in a managed population such as the honey bee, will be more dependent on beekeeping activities.

As was true with detoxification, selection for alleles associated with immunity makes biological sense. Honey bees face a multitude of enemies which, in addition to pesticides, are recognized as key contributors to worldwide population declines (reviewed by vanEngelsdorp & Meixner 2010; Potts *et al.* 2010). While honey bees have evolved both group (e.g. grooming, nest hygiene, necrophoric behaviour, behavioural fever) and individual (mechanical, physiological and immune responses) defence mechanisms, their health has been challenged in unprecedented ways by an increasing number of emerging enemies including parasites, pathogens and predators. Among the most significant are the parasitic mite *Varroa destructor*, the gram-positive bacterium *Paenibacillus larvae* and the microsporidia *Nosema apis* and *N. ceranae* (vanEngelsdorp & Meixner 2010). The latter has been particularly worrisome in Spain, where it has been reported as a highly virulent pathogen and a potential culprit in colony collapse disorder (Higes *et al.* 2009, 2010), although epidemiological studies in the US (Cox-Foster *et al.* 2007; vanEngelsdorp *et al.* 2009) and Germany (Gisder *et al.* 2010) failed to associate this emerging enemy to colony losses. While the impact of *V. destructor* (on its own and as a vector of viruses) has been alleviated by in-hive acaricides (although mite-resistance has been reported for an increasing number of chemicals), colonies are rarely treated against the pathogens *P. larvae* and *Nosema* spp. because antibiotics are interdicted in the European Union due to honey contamination. Accordingly, Iberian honey bees have likely been under pathogen-driven selective pressure, although only two strong outliers, both mapped to genes putatively involved in phagocytosis immune response, were identified in this study. In *D. melanogaster*, genes implicated in phagocytosis have been found to have an important role in deterring bacteria (Kurucz *et al.* 2007) and fungi (Jin *et al.* 2008). While little is known about the importance of phagocytosis in honey bees (Evans & Spivak 2010), gene-expression studies have reported differential regulation of *NIMC2* or *DSCAM* genes in colonies exposed to different parasites and pathogens (Evans 2006; Navajas *et al.* 2008; Bull *et al.* 2012; Nazzi *et al.* 2012) suggesting that these genes might participate in the immune response.

The SNP outliers identified in this study are only a small fraction of the total SNP variation in the honey bee. As such, they are markers of selection, and unlikely to be direct targets of selection. Many of the outlier SNPs were located in introns and UTRs and likely have little or no impact on the polypeptide, although these same regions may be involved in regulation or alternative splicing (Ladd & Cooper 2002). Those outlier SNPs that were located in exons were all silent substitutions,

which have widely been assumed to have no effect on protein fitness, although comparison of all dipteran and hymenopteran sequenced genomes, shows that codon usage bias is highest in the honey bee (Behura & Severson 2012) and is associated with differential translation rates, mRNA stability and modification of protein structure and activity (Tsai *et al.* 2008). A few SNPs were located outside coding regions. Accordingly, it is likely that many SNPs, if not all, have hitchhiked with linked polymorphisms that are the causative mutations under selection.

Differentiation of neutral and directional outlier loci

Partition of neutral variation was negligible across the Iberian honey bee range, barely reflecting the steep north–south maternal cline formed by the two highly divergent western European (M) and African (A) lineages extensively reported in the literature (Garnery *et al.* 1995; Franck *et al.* 1998; Arias *et al.* 2006; Miguel *et al.* 2007; Cánovas *et al.* 2008; Pinto *et al.* 2013) and further confirmed in this study (Fig. S1, Supporting information). A pattern of virtually no differentiation was also revealed by geometric morphometric data (Miguel *et al.* 2011) and microsatellites (Franck *et al.* 1998; Garnery *et al.* 1998b; Cánovas *et al.* 2011) but not by traditional morphometry (Ruttner *et al.* 1978; Cornuet & Fresnaye 1989) and the allozyme malate dehydrogenase (Nielsen *et al.* 1994; Smith & Glenn 1995; Arias *et al.* 2006), which have been reported to exhibit a latitudinal gradient along the Mediterranean coast of the Iberian Peninsula.

Contrasting with neutral loci, geographical partitioning recovered by directional outliers argues for selection as an underlying force shaping a latitudinal gradient, as suggested by others (Franck *et al.* 1998; Garnery *et al.* 1998a,b). The spatial analysis identified precipitation, and to a lesser extent temperature as associated with the north–south (latitudinal) selection (Table 1). Precipitation was associated along with latitude for nine outlier SNPs, while mean and maximum temperatures were also associated with the precipitation and latitude for four of these nine. What was not expected was that several directional outliers were found even more strongly correlated with longitude than latitude (see AMOVA and matSAM analysis). Further, while matSAM identified precipitation associated with both longitude and latitude, insolation was identified as uniquely associated with longitude. Insolation is thus identified as a novel parameter shaping population structure east to west. The observation of east–west outlier structuring further adds to the complexity of Iberian honey bees suggesting that the Atlantic portion of the Iberian Peninsula harbours an additional important component of the Iberian honey bee diversity.

Concluding remarks

In this study, genome-wide scans using 383 polymorphic SNP loci were conducted in searching for the footprints of selection in the environmentally heterogeneous Iberian honey bee range. A combination of methods was employed leading to the detection of several SNPs marking genomic regions that are promising candidates for adaptation. SNP loci exhibiting strong outlier behaviour were cross-validated by at least three conceptually different F_{ST}-based methods, and a subset was found to be associated with environmental variables that may be causal or correlated selection pressures.

Further support for the role of selection in shaping variation in Iberian honey bees was provided by the functional relevance of the genes carrying outlier SNPs. This is particularly compelling when independent SNPs are mapped to genes that are functionally related. While genome-wide scans provide a powerful way of highlighting candidate genes for selection, experimental support for selection from functional and expression studies (Schluter *et al.* 2010; Renaut *et al.* 2011; Riveron *et al.* 2013) and also indirect evidence from sequence variation analysis (Low *et al.* 2007; Wood *et al.* 2008; Kent *et al.* 2011; Oliveira *et al.* 2012) are ultimately required to make causal inferences about the molecular basis of adaptation of Iberian honey bees. Nevertheless, the approach pursued here enabled, for the first time, a more rigorous consideration of selection as the underlying cause of diversity patterns in Iberian honey bees, whereas in previous studies (Franck *et al.* 1998; Garnery *et al.* 1998a,b), evocation of this evolutionary force to explain clinal variation in the Iberian Peninsula was speculative. Furthermore, our findings represent an important first step towards the identification of polymorphisms implicated in local adaptation and possibly in response to recent human-mediated environmental changes.

Acknowledgements

We are deeply indebted to numerous beekeepers and technicians of beekeepers' associations and of government institutions for inestimable assistance with the honey bee collection. Special appreciation is due to Antonio Gomez Pajuelo for providing numerous contacts of beekeepers in Spain; to Andreia Brandão, Inês Moura, and Margarida Neto for helping in the field work; to Danny Weaver for permission to use the OPA; to Colette Abbey for DNA extractions and running GoldenGate assays; to Clare Gill for assistance with the GenomeStudio software and for providing the facilities at Texas A&M University for SNP genotyping; to Miguel Vaz-Pinto and João Paulo Castro for GIS and GPS assistance; to José Rufino for providing computer power for running the simulations. GoldenGate SNP data were produced with support from the Whole Systems

Genomics Initiative at Texas A&M University in cooperation with Texas A&M AgriLife Research. Five reviewers made constructive and insightful comments on an earlier draft of the manuscript. Julio Chávez-Galarza was supported by Fundação para a Ciência e Tecnologia through the PhD scholarship SFRH/BD/68682/2010. Irene Muñoz was supported by the Spanish Ministry of Education, Culture and Sports. Financial support for this research was provided by Fundação para a Ciência e Tecnologia and COMPETE/QREN/EU through the project PTDC/BIA-BEC/099640/2008 to M. A. Pinto.

References

Ai H, Huang L, Ren J (2013) Genetic diversity, linkage disequilibrium and selection signatures in chinese and Western pigs revealed by genome-wide SNP markers. *PLoS One*, **8**, e56001.

Allendorf FW, Luikart G (2007) *Conservation and the Genetics of Populations*, 1st edn. Blackwell Publishing, Malden.

Antão T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: a workbench to detect molecular adaptation based on a F_{ST}-outlier method. *BMC Bioinformatics*, **9**, 323.

Arias MC, Rinderer TE, Sheppard WS (2006) Further characterization of honey bees from the Iberian Peninsula by allozyme, morphometric and mtDNA haplotype analyses. *Journal of Apicultural Research*, **45**, 188–196.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**, 969–980.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, **263**, 1619–1626.

Behura SK, Severson DW (2012) Comparative analysis of codon usage bias and codon context patterns between dipteran and hymenopteran sequenced genomes. *PLoS One*, **7**, e43111.

Belyaeva OV, Lee SA, Kolupaev OV, Kedishvili NY (2009) Identification and characterization of retinoid-active short-chain dehydrogenases/reductases in *Drosophila melanogaster*. *Biochimica et Biophysica Acta-General Subjects*, **1790**, 1266–1273.

Beye M, Gattermeier I, Hasselmann M *et al.* (2006) Exceptionally high levels of recombination across the honey bee genome. *Genome Research*, **16**, 1339–1344.

Blacquière T, Smagghe G, van Gestel CAM, Mommaerts V (2012) Neonicotinoids in bees: a review on concentrations, side-effects and risk assessment. *Ecotoxicology*, **21**, 1581.

Bonin A, Taberlet P, Miaud C, Pompanon F (2006) Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Molecular Biology and Evolution*, **23**, 773–783.

Buchon N, Broderick NA, Poidevin M, Pradervand S, Lemaitre B (2009) *Drosophila* intestinal response to bacterial infection: activation of host defense and stem cell proliferation. *Cell Host & Microbe*, **5**, 200–211.

Buckley J, Butlin RK, Bridle JR (2012) Evidence for evolutionary change associated with the recent range expansion of the British butterfly, *Aricia agestis*, in response to climate change. *Molecular Ecology*, **21**, 267–280.

Bull DL, Whitten CJ (1972) Factors influencing organophosphorus insecticide resistance in tobacco budworms. *Journal of Agricultural and Food Chemistry*, **20**, 561–564.

Bull JC, Ryabov EV, Prince G *et al.* (2012) A strong immune response in young adult honey bees masks their increased susceptibility to infection compared to older bees. *PLoS Pathogens*, **8**, e1003083.

Bunce JA, Isbell LA, Neitz M *et al.* (2011) Characterization of opsin gene alleles affecting color vision in a wild population of titi monkeys (*Callicebus brunneus*). *American Journal of Primatology*, **73**, 189–196.

Burrill RM, Dietz A (1981) The response of honey bees to variations in solar radiation and temperature. *Apidologie*, **12**, 319–328.

Cafferty P, Yu L, Long H, Rao Y (2006) Semaphorin-1a functions as a guidance receptor in the *Drosophila* visual system. *Journal of Neuroscience*, **26**, 3999–4003.

Camps M, Herman A, Loh E, Loeb LA (2007) Genetic constraints on protein evolution. *Critical Reviews in Biochemistry and Molecular Biology*, **42**, 313–326.

Cánovas F, De la Rúa P, Serrano J, Galián J (2008) Geographical patterns of mitochondrial DNA variation in *Apis mellifera iberiensis* (Hymenoptera: Apidae). *Journal of Zoological Systematics and Evolutionary Research*, **46**, 24–30.

Cánovas F, De la Rúa P, Serrano J, Galián J (2011) Microsatellite variability reveals beekeeping influences on Iberian honey bee populations. *Apidologie*, **42**, 235–251.

Claudianos C, Ranson H, Johnson RM *et al.* (2006) A deficit of detoxification enzymes: pesticide sensitivity and environmental response in the honey bee. *Insect Molecular Biology*, **15**, 615–636.

Cooke GM, Chao NL, Beheregaray LB (2012) Divergent natural selection with gene flow along major environmental gradients in Amazonia: insights from genome scans, population genetics and phylogeography of the characin fish *Triportheus albus*. *Molecular Ecology*, **21**, 2410–2427.

Coop G, Witonsky D, Di Rienzo A, Pritchard JK (2010) Using environmental correlations to identify loci underlying local adaptation. *Genetics*, **185**, 1411–1423.

Cornuet JM, Fresnaye J (1989) Biometrical study of honey bee populations from Spain and Portugal. *Apidologie*, **20**, 93–101.

Cox-Foster DL, Conlan S, Holmes EC *et al.* (2007) A metagenomic survey of microbes in honey bee colony collapse disorder. *Science*, **318**, 283–287.

De Jong MA, Collins S, Beldade P, Brakefield PM, Zwaan BJ (2013) Footprints of selection in wild populations of *Bicyclus anynana* along a latitudinal cline. *Molecular Ecology*, **22**, 341–353.

Dietemann V, Pirk CWW, Crewe R (2009) Is there a need for conservation of honey bees in Africa? *Apidologie*, **40**, 285–295.

Dussaubat C, Brunet JL, Higes M *et al.* (2012) Gut pathology and responses to the microsporidium *Nosema ceranae* in the honey bee *Apis mellifera*. *PLoS One*, **7**, e37017.

Earl JB, Britt SG (2006) Expression of *Drosophila* rhodopsins during photoreceptor cell differentiation: insights into R7 and R8 cell subtype commitment. *Gene Expression Patterns*, **6**, 687–694.

Engel MS (1999) The taxonomy of recent and fossil honey bees (Hymenoptera: Apidae; *Apis*). *Journal of Hymenoptera Research*, **8**, 165–196.

vanEngelsdorp D, Meixner MD (2010) A historical review of managed honey bee populations in Europe and the United States and the factors that may affect them. *Journal of Invertebrate Pathology*, **103**, S80–S95.

vanEngelsdorp D, Evans JD, Saegerman C *et al.* (2009) Colony collapse disorder: a descriptive study. *PLoS One*, **4**, e6481.

Evans JD (2006) Beepath: an ordered quantitative-PCR array for exploring honey bee immunity and disease. *Journal of Invertebrate Pathology*, **93**, 135–139.

Evans JD, Spivak M (2010) Socialized medicine: individual and communal disease barriers in honey bees. *Journal of Invertebrate Pathology*, **103**, S62–S72.

Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity (Edinb)*, **103**, 285–298.

Feyereisen R (1999) Insect P450 Enzymes. *Annual Review of Entomology*, **44**, 507–533.

Ffrench-Constant RH, Daborn PJ, Le Goff G (2004) The genetics and genomics of insecticide resistance. *Trends in Genetics*, **20**, 163–170.

Foll M, Gaggiotti O (2006) Identifying the environmental factors that determine the genetic structure of populations. *Genetics*, **174**, 875–891.

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, **180**, 977–993.

Franck P, Garnery L, Solignac M, Cornuet JM (1998) The origin of west European subspecies of honey bees (*Apis mellifera*): new insights from microsatellite and mitochondrial data. *Evolution*, **52**, 1119–1134.

Garnery L, Mosshine EH, Oldroyd BP, Cornuet JM (1995) Mitochondrial DNA variation in Moroccan and Spanish honey bee populations. *Molecular Ecology*, **4**, 465–471.

Garnery L, Franck P, Baudry E *et al.* (1998a) Genetic diversity of the west European honey bee (*Apis mellifera mellifera* and *A. m. iberica*). I. Mitochondrial DNA. *Genetics Selection Evolution*, **30**, S31–S47.

Garnery L, Franck P, Baudry E *et al.* (1998b) Genetic diversity of the west European honey bee (*Apis mellifera mellifera* and *A. m. iberica*). II. Microsatellite loci. *Genetics Selection Evolution*, **30**, S49–S74.

Gisder S, Hedtke K, Mockel N *et al.* (2010) Five-year cohort study of *Nosema spp.* in Germany: does climate shape virulence and assertiveness of *Nosema ceranae*? *Applied and Environmental Microbiology*, **76**, 3032–3038.

Godenschwege TA, Hu HL, Shan-Crofts X, Goodman CS, Murphey RK (2002) Bi-directional signaling by Semaphorin 1a during central synapse formation in *Drosophila*. *Nature Neuroscience*, **5**, 1294–1301.

Gomez-Uchida D, Seeb JE, Smith MJ *et al.* (2011) Single nucleotide polymorphisms unravel hierarchical divergence and signatures of selection among Alaskan sockeye salmon (*Oncorhynchus nerka*) populations. *BMC Evolutionary Biology*, **11**, 48.

Hardstone MC, Scott JG (2010) Is *Apis mellifera* more sensitive to insecticides than other insects? *Pest Management Science*, **66**, 1171–1180.

Helyar SJ, Hemmer-Hansen J, Bekkevold D *et al.* (2011) Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Molecular Ecology Resources*, **11**, 123–136.

Heymann Y, Steenmans C, Croissille G, Bossard M (1994) *CORINE Land Cover*. Office for Official Publications of the European Communities, Luxembourg.

Higes M, Martin-Hernandez R, Garrido-Bailon E *et al.* (2009) Honeybee colony collapse due to *Nosema ceranae* in professional apiaries. *Environmental Microbiology Reports*, **1**, 110–113.

Higes M, Martin-Hernandez R, Meana A (2010) *Nosema ceranae* in Europe: an emergent type C nosemosis. *Apidologie*, **41**, 375–392.

Hollox EJ, Armour JAL (2008) Directional and balancing selection in human beta-defensins. *BMC Evolutionary Biology*, **8**, 113.

Jaffe R, Dietemann V, Allsopp MH *et al.* (2010) Estimating the density of honey bee colonies across their natural range to fill the gap in pollinator decline censuses. *Conservation Biology*, **24**, 583–593.

Jiggins FM, Hurst GDD (2003) The evolution of parasite recognition genes in the innate immune system: purifying selection on *Drosophila melanogaster* peptidoglycan recognition proteins. *Journal of Molecular Evolution*, **57**, 598–605.

Jin LH, Shim J, Yoon JS *et al.* (2008) Identification and functional analysis of antifungal immune response genes in *Drosophila*. *PLoS Pathogens*, **4**, e1000168.

Johnson RM, Wen Z, Schuler MA, Berenbaum MR (2006) Mediation of pyrethroid insecticide toxicity to honey bees (Hymenoptera: Apidae) by cytochrome P450 monooxygenases. *Journal of Economic Entomology*, **99**, 1046–1050.

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.

Joost S, Bonin A, Bruford MW *et al.* (2007) A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Molecular Ecology*, **16**, 3955–3969.

Joost S, Kalbermatten M, Bonin A (2008) Spatial analysis method(SAM): a software tool combining molecular and environmental data to identify candidate loci for selection. *Molecular Ecology Resources*, **8**, 957–960.

Kent CF, Issa A, Bunting AC, Zayed A (2011) Adaptive evolution of a key gene affecting queen and worker traits in the honey bee, *Apis mellifera*. *Molecular Ecology*, **20**, 5226–5235.

Kim TS, Maeda A, Maeda T *et al.* (2005) Delayed dark adaptation in 11-cis-retinol dehydrogenase-deficient mice—a role of RDH11 in visual processes in vivo. *Journal of Biological Chemistry*, **280**, 8694–8704.

Kinel-Tahan Y, Weiss H, Dgany O, Levine A, Wides R (2007) *Drosophila* odz gene is required for multiple cell types in the compound retina. *Developmental Dynamics*, **236**, 2541–2554.

Kurucz E, Markus R, Zsamboki J *et al.* (2007) Nimrod, a putative phagocytosis receptor with EGF repeats in *Drosophila* plasmatocytes. *Current Biology*, **17**, 649–654.

Ladd AN, Cooper TA (2002) Finding signals that regulate alternative splicing in the postgenomic era. *Genome Biology*, **3**, reviews0008.

Lehtonen PK, Laaksonen T, Artemyev AV *et al.* (2012) Candidate genes for colour and vision exhibit signals of selection across the pied flycatcher (*Ficedula hypoleuca*) breeding range. *Heredity*, **108**, 431–440.

Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, **74**, 175–195.

Li X, Schuler MA, Berenbaum MR (2007) Molecular mechanisms of metabolic resistance to synthetic and natural xenobiotics. *Annual Review of Entomology*, **52**, 231–253.

Limborg MT, Helyar SJ, de Bruyn M *et al.* (2012) Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (*Clupea harengus*). *Molecular Ecology*, **21**, 3686–3703.

Low WY, Ng HL, Morton CJ *et al.* (2007) Molecular evolution of glutathione S-transferases in the genus *Drosophila*. *Genetics*, **177**, 1363–1375.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews. Genetics*, **4**, 981–994.

Manel S, Poncet BN, Legendre P, Gugerli F, Holderegger R (2010) Common factors drive adaptive genetic variation at different spatial scales in *Arabis alpina*. *Molecular Ecology*, **19**, 3824–3835.

Mao W, Rupasinghe SG, Johnson RM *et al.* (2009) Quercetin-metabolizing CYP6AS enzymes of the pollinator *Apis mellifera* (Hymenoptera: Apidae). *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, **154**, 427–434.

Mao W, Schuler MA, Berenbaum MR (2011) CYP9Q-mediated detoxification of acaricides in the honey bee (*Apis mellifera*). *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 12657–12662.

Meixner MD, Leta MA, Koeniger N, Fuchs S (2011) The honey bees of Ethiopia represent a new subspecies of *Apis mellifera*–*Apis mellifera simensis* n. ssp. *Apidologie*, **42**, 425–437.

Menzel R, Müller U (1996) Learning and memory in honey bees: from behavior to neural substrates. *Annual Review of Neuroscience*, **19**, 379–404.

Miguel I, Iriondo M, Garnery L, Sheppard WS, Estonba A (2007) Gene flow within the M evolutionary lineage of *Apis mellifera*: role of the Pyrenees, isolation by distance and post-glacial re-colonization routes in the western Europe. *Apidologie*, **38**, 141–155.

Miguel I, Baylac M, Iriondo M *et al.* (2011) Both geometric morphometric and microsatellite data consistently support the differentiation of the *Apis mellifera* M evolutionary branch. *Apidologie*, **42**, 150–161.

Moritz RFA, Kraus FB, Kryger P, Crewe RM (2007) The size of wild honey bee populations (*Apis mellifera*) and its implications for the conservation of honey bees. *Journal of Insect Conservation*, **11**, 391–397.

Narum SR, Hess JE (2011) Comparison of F_{ST} outlier tests for SNP loci under selection. *Molecular Ecology Resources*, **11**, 184–194.

Navajas M, Migeon A, Alaux C *et al.* (2008) Differential gene expression of the honey bee *Apis mellifera* associated with *Varroa destructor* infection. *BMC Genomics*, **9**, 301.

Nazzi F, Brown SP, Annoscia D *et al.* (2012) Synergistic parasite–pathogen interactions mediated by host immunity can drive the collapse of honey bee colonies. *PLoS Pathogens*, **8**, e1002735.

Nielsen R (2005) Molecular signatures of natural selection. *Annual Review of Genetics*, **39**, 197–218.

Nielsen D, Page RE Jr, Crosland MW (1994) Clinal variation and selection of MDH allozymes in honey bee populations. *Experientia*, **50**, 867–871.

Nielsen EE, Hemmer-Hansen J, Poulsen NA *et al.* (2009) Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). *BMC Evolutionary Biology*, **9**, 276.

Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, **18**, 375–402.

Nunes VL, Beaumont MA, Butlin RK, Paulo OS (2011) Multiple approaches to detect outliers in a genome scan for selection in ocellated lizards (*Lacerta lepida*) along an environmental gradient. *Molecular Ecology*, **20**, 193–205.

Oliveira TYK, Harris EE, Meyer D, Jue CK, Silva WA (2012) Molecular evolution of a malaria resistance gene (DARC) in primates. *Immunogenetics*, **64**, 497–505.

Osborne JL (2012) Ecology: Bumblebees and pesticides. *Nature*, **491**, 43–45.

Pérez-Figueroa A, Garcia-Pereira MJ, Saura M, Rolan-Alvarez E, Caballero A (2010) Comparing three different methods to detect selective loci using dominant markers. *Journal of Evolutionary Biology*, **23**, 2267–2276.

Pilling ED, Bromleychallenor KAC, Walker CH, Jepson PC (1995) Mechanism of synergism between the pyrethroid insecticide lambda-cyhalothrin and the imidazole fungicide prochloraz, in the honey bee (*Apis mellifera* L.). *Pesticide Biochemistry and Physiology*, **51**, 1–11.

Pinto MA, Henriques D, Neto M *et al.* (2013) Maternal diversity patterns of Ibero-Atlantic populations reveal further complexity of Iberian honey bees. *Apidologie*, **44**, 430–439.

Potts SG, Biesmeijer JC, Kremen C *et al.* (2010) Global pollinator declines: trends, impacts and drivers. *Trends in Ecology & Evolution*, **25**, 345–353.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Prunier J, Laroche J, Beaulieu J, Bousquet J (2011) Scanning the genome for gene SNPs related to climate adaptation and estimating selection at the molecular level in boreal black spruce. *Molecular Ecology*, **20**, 1702–1716.

Renaut S, Nolte AW, Rogers SM, Derome N, Bernatchez L (2011) SNP signatures of selection on standing genetic variation and their association with adaptive phenotypes along gradients of ecological speciation in lake whitefish species pairs (*Coregonus spp.*). *Molecular Ecology*, **20**, 545–559.

Riveron JM, Irving H, Ndula M *et al.* (2013) Directionally selected cytochrome P450 alleles are driving the spread of pyrethroid resistance in the major malaria vector *Anopheles funestus*. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 252–257.

Ruttner F (1988) *Biogeography and Taxonomy of Honey Bees*. Springer, Berlin.

Ruttner F, Tassencourt L, Louveaux J (1978) Biometrical-statistical analysis of the geographic variability of *Apis mellifera* L. *Apidologie*, **9**, 363–381.

Sackton TB, Lazzaro BP, Schlenke TA *et al.* (2007) Dynamic evolution of the innate immune system in *Drosophila*. *Nature Genetics*, **39**, 1461–1468.

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning*, 2nd edn. Cold Spring Harbor Laboratory Press, New York.
Schlötterer C (2002) A microsatellite-based multilocus screen for the identification of local selective sweeps. *Genetics*, **160**, 753–763.
Schluter D, Marchinko KB, Barrett RDH, Rogers SM (2010) Natural selection and the genetics of adaptation in three-spine stickleback. *Philosophical Transactions of the Royal Society Series B: Biological Sciences*, **365**, 2479–2486.
Scott JG, Wen ZM (2001) Cytochromes P450 of insects: the tip of the iceberg. *Pest Management Science*, **57**, 958–967.
Sheppard WS, Meixner MD (2003) *Apis mellifera pomonella*, a new honey bee subspecies from Central Asia. *Apidologie*, **34**, 367–375.
Shikano T, Ramadevi J, Merilä J (2010) Identification of local- and habitat-dependent selection: scanning functionally important genes in nine-spined sticklebacks (*Pungitius pungitius*). *Molecular Biology and Evolution*, **27**, 2775–2789.
Shimada Y, Shikano T, Merilä J (2011) A high incidence of selection on physiologically important genes in the three-spined stickleback, *Gasterosteus aculeatus*. *Molecular Biology and Evolution*, **28**, 181–193.
Smith DR, Glenn TC (1995) Allozyme polymorphisms in spanish honey bees (*Apis mellifera iberica*). *Journal of Heredity*, **86**, 12–16.
Smith DR, Palopoli MF, Taylor BR *et al.* (1991) Geographical overlap of two mitochondrial genomes in Spanish honey bees (*Apis mellifera iberica*). *Journal of Heredity*, **82**, 96–100.
Somogyi K, Sipos B, Penzes Z *et al.* (2008) Evolution of genes and repeats in the Nimrod superfamily. *Molecular Biology and Evolution*, **25**, 2337–2347.
Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology*, **14**, 671–688.
Thioulouse J, Chessel D, Doledec S, Olivier JM (1997) ADE-4: a multivariate analysis and graphical display software. *Statistics and Computing*, **7**, 75–83.
Townson SM, Chang BSW, Salcedo E *et al.* (1998) Honeybee blue-and ultraviolet-sensitive opsins: cloning, heterologous expression in *Drosophila*, and physiological characterization. *Journal of Neuroscience*, **18**, 2412–2422.
Tsai CJ, Sauna ZE, Kimchi-Sarfaty C *et al.* (2008) Synonymous mutations and ribosome stalling can lead to altered folding pathways and distinct minima. *Journal of Molecular Biology*, **383**, 281–291.
Tsumura Y, Uchiyama K, Moriguchi Y, Ueno S, Ihara-Ujino T (2012) Genome scanning for detecting adaptive genes along environmental gradients in the Japanese conifer, *Cryptomeria japonica*. *Heredity*, **109**, 349–360.
Vasemägi A, Primmer CR (2005) Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Molecular Ecology*, **14**, 3623–3642.
Vitalis R, Dawson K, Boursot P (2001) Interpretation of variation across marker loci as evidence of selection. *Genetics*, **158**, 1811–1823.
Wang Q, Hasan G, Pikielny CW (1999) Preferential expression of biotransformation enzymes in the olfactory organs of *Drosophila melanogaster*, the antennae. *Journal of Biological Chemistry*, **274**, 10309–10315.
Wang XY, Wang T, Jiao YC, von Lintig J, Montell C (2010) Requirement for an enzymatic visual cycle in *Drosophila*. *Current Biology*, **20**, 93–102.
Wang T, Chen GP, Zan QJ, Wang CB, Su YJ (2012) AFLP genome scan to detect genetic structure and candidate loci under selection for local adaptation of the invasive weed *Mikania micrantha*. *PLoS One*, **7**, e41310.
Watson FL, Puttmann-Holgado R, Thomas F *et al.* (2005) Extensive diversity of Ig-superfamily proteins in the immune system of insects. *Science*, **309**, 1874–1878.
Whitfield CW, Behura SK, Berlocher SH *et al.* (2006) Thrice out of Africa: ancient and recent expansions of the honey bee, *Apis mellifera*. *Science*, **314**, 642–645.
Wilding CS, Butlin RK, Grahame J (2001) Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. *Journal of Evolutionary Biology*, **14**, 611–619.
Winston M (1987) *The Biology of the Honey Bee*. Harvard University Press, London.
Wood HM, Grahame JW, Humphray S, Rogers J, Butlin RK (2008) Sequence differentiation in regions identified by a genome scan for local adaptation. *Molecular Ecology*, **17**, 3123–3135.
Yu SJ (2002) Biochemical characteristics of microsomal and cytosolic glutathione S-transferases in larvae of the fall armyworm, *Spodoptera frugiperda* (J. E. Smith). *Pesticide Biochemistry and Physiology*, **72**, 100–110.
Zayed A, Whitfield CW (2008) A genome-wide signature of positive selection in ancient and recent invasive expansions of the honey bee *Apis mellifera*. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 3421–3426.

M.A.P., J.C.P., and J.S.J. designed the study. J.C.-G. performed all the analyses and searched for the GO information. D.H. assisted J.C.-G. with files' preparation, Structure and AMOVA analyses. M.A.P., I.M. and P.D.L.R. coordinated sampling across Iberia and performed sampling across the Atlantic transect. J.C.A. produced the sampling maps, conducted the land use-allele frequency analysis, and assisted with GIS. J.S.J. coordinated SNP genotyping. M.A.P. validated SNP genotypes. J.C.-G, M.A.P., J.S.J and D.H. interpreted the results. J.C.-G. and M.A.P. wrote the paper with input from the other authors.

Data accessibility

SNP genotypes for the 711 Iberian honey bee individuals: DRYAD entry doi:10.5061/dryad.1kk2k.

Geographical coordinates of sampling locations and environmental data for the 711 Iberian honey bee individuals: DRYAD entry doi:10.5061/dryad.1kk2k.

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Individual bar charts indicate allele frequencies of mtDNA (M-lineage haplotypes), the strongest directional outliers (from ahb1245 to est11018), the strongest balancing outlier (AMB-00963630), and a representative neutral locus, in each of the 23 sampling sites.

Fig. S2 Graphical representation of the five groupings used in the AMOVA (see Table 2).

Fig. S3 Population structure and admixture levels obtained with the software Structure based on 1075 SNP loci.

Fig. S4 Principal component analysis (PCA) based on 1075 SNP loci. PCA was performed on a normalized matrix of individuals vs. SNP loci.

Fig. S5 (a) LOSITAN, (b) Arlequin, (c) BAYESFST and (d) BAYESCAN plots showing loci under selection.

Table S1 Genomic information obtained from NCBI (http://www.ncbi.nlm.nih.gov/), BEEBASE (http://hymenopteragenome.org/beebase), and FLYBASE (www.flybase.org) for the remaining directional and balancing outlier SNP loci as detected by two and one F_{ST}-based method (P-values ≤ 0.05), and for SNP loci that were associated with at least one environmental variable identified by the spatial analysis method (matSAM).

Table S2 Environmental variables associated with SNP loci (CI $\geq$99% with Bonferroni correction), as detected by the spatial analysis method (matSAM).

Table S3 Accession numbers of genes marked by SNPs that were detected by at least one of the five methods employed listed in Table 1 and Table S1.

MOLECULAR ECOLOGY

Molecular Ecology (2015) **24**, 2973–2992 doi: 10.1111/mec.13223

Revisiting the Iberian honey bee (*Apis mellifera iberiensis*) contact zone: maternal and genome-wide nuclear variations provide support for secondary contact from historical refugia

JULIO CHÁVEZ-GALARZA,*† DORA HENRIQUES,*† J. SPENCER JOHNSTON,‡ MIGUEL CARNEIRO,§ JOSÉ RUFINO,¶ JOHN C. PATTON** and M. ALICE PINTO*

**Mountain Research Centre (CIMO), Polytechnic Institute of Bragança, Campus de Sta. Apolónia, Apartado 1172, 5301-855 Bragança, Portugal, †Centre of Molecular and Environmental Biology (CBMA), University of Minho, Campus de Gualtar 4710-057 Braga, Portugal, ‡Department of Entomology, Texas A&M University, College Station, TX 77843-2475, USA, §CIBIO/InBIO, Research Center in Biodiversity and Genetic Resources, University of Porto, Campus Agrário de Vairão, 4485-661 Vairão, Portugal, ¶Polytechnic Institute of Bragança, 5301-857 Bragança, Portugal, **Department of Forestry and Natural Resources, Purdue University, 715 W State St., West Lafayette, IN 4797-2061, USA*

Abstract

Dissecting diversity patterns of organisms endemic to Iberia has been truly challenging for a variety of taxa, and the Iberian honey bee is no exception. Surveys of genetic variation in the Iberian honey bee are among the most extensive for any honey bee subspecies. From these, differential and complex patterns of diversity have emerged, which have yet to be fully resolved. Here, we used a genome-wide data set of 309 neutrally tested single nucleotide polymorphisms (SNPs), scattered across the 16 honey bee chromosomes, which were genotyped in 711 haploid males. These SNPs were analysed along with an intergenic locus of the mtDNA, to reveal historical patterns of population structure across the entire range of the Iberian honey bee. Overall, patterns of population structure inferred from nuclear loci by multiple clustering approaches and geographic cline analysis were consistent with two major clusters forming a well-defined cline that bisects Iberia along a northeastern–southwestern axis, a pattern that remarkably parallels that of the mtDNA. While a mechanism of primary intergradation or isolation by distance could explain the observed clinal variation, our results are more consistent with an alternative model of secondary contact between divergent populations previously isolated in glacial refugia, as proposed for a growing list of other Iberian taxa. Despite current intense honey bee management, human-mediated processes have seemingly played a minor role in shaping Iberian honey bee genetic structure. This study highlights the complexity of the Iberian honey bee patterns and reinforces the importance of Iberia as a reservoir of *Apis mellifera* diversity.

Keywords: *Apis mellifera iberiensis*, geographic cline analysis, honey bee, Iberia, secondary contact, SNPs, sPCA, structure

Received 3 August 2014; revision received 16 April 2015; accepted 21 April 2015

Introduction

Clinal patterns in gene frequencies can be generated by random genetic drift under an isolation-by-distance scenario. Alternatively, clinal variation may be shaped by selection acting within a continuous population (primary intergradation) or, more frequently, may originate from contact between populations that diverged in allopatry (secondary contact). Distinguishing primary intergradation from secondary contact can, however, be a difficult undertaking because both processes may

Correspondence: M. Alice Pinto, Fax: +351273325405; E-mail: apinto@ipb.pt

generate similar patterns of genetic variation (Endler 1977; Barton & Hewitt 1985, 1989). Population genomics provides a suitable framework in which to more effectively unravel such levels of complexity. In population genomics, outlier tests are applied to genome-wide sampling of multiple populations to dissect out adaptive variation, leaving a background of neutral and near-neutral variation (Luikart *et al.* 2003). Cline analysis can then help reveal whether dissected patterns of variation originated from secondary contact or primary intergradation. If multiple coincident clines are identified (Endler 1977; Barton & Hewitt 1985) and these clines reflect changes in neutral loci, there is strong support for recent secondary contact (Durrett *et al.* 2000). Unless many independent loci respond similarly to a single environmental gradient or mosaic, clinal patterns of neutral variation and multiple coincident clines are not expected when primary intergradation is the leading process shaping variation (Durrett *et al.* 2000).

The Iberian Peninsula provides one of the most interesting settings in Europe for studying contact zones. High geological, physiographical and climatic complexity and diversity, together with isolation from Europe and proximity to Africa (especially at the Strait of Gibraltar), made this southernmost European region an important refuge during the Quaternary glaciations (reviewed by Hewitt 2000) and a bridge, for the more vagile organisms, between the two continents (Carranza *et al.* 2004; Cosson *et al.* 2005; Guillaumet *et al.* 2006; Whitfield *et al.* 2006; Wallberg *et al.* 2014). These features made Iberia not only a place of divergence during periods of isolation but also a contact zone during periods of expansion as reported for a wide array of plant and animal taxa (extensively reviewed by Weiss & Ferrand 2007), including the focal organism of this study: the Iberian honey bee, *Apis mellifera iberiensis*.

Disentangling diversity patterns in populations that have possibly experienced recurrent cycles of contraction, expansion, admixture, and adaptation, typical of long-term glacial refugia, is a challenging endeavour. Contemporary human-mediated processes, which in the case of the honey bee may involve movement of colonies within (transhumance) and between lineages (introduction of commercial queens), selective breeding, and accidental introductions of exotic pests and diseases, may further complicate this effort by erasing or obscuring the genetic signatures imprinted by evolutionary and demographic processes. Fortunately, however, Iberia is the best-studied refugial area in Europe, and common patterns are emerging from comparative phylogeography (Gómez & Lunt 2007) that are of great assistance in elucidating patterns exhibited by the Iberian honey bee. Additionally, the honey bee genome has been sequenced and a SNP panel is available for conducting genome-wide sampling of multiple populations (Whitfield *et al.* 2006; Chávez-Galarza *et al.* 2013).

The honey bee native range spans Africa, Europe and the Middle East where it evolved into 30 subspecies (Ruttner 1988; Engel 1999; Sheppard & Meixner 2003; Meixner *et al.* 2011). This vast amount of variation has been grouped into four (western European, M; eastern European, C; African, A; Middle Eastern, O) largely parapatric evolutionary lineages (Ruttner 1988; Garnery *et al.* 1992; Whitfield *et al.* 2006; Wallberg *et al.* 2014), with contact zones identified in Italy (Franck *et al.* 2000), Turkey (Kandemir *et al.* 2006), Libya (Shaibi *et al.* 2009), and Iberia (Smith *et al.* 1991; Franck *et al.* 1998; Garnery *et al.* 1998a; Arias *et al.* 2006; Miguel *et al.* 2007; Cánovas *et al.* 2008). Among these, the Iberian contact zone formed by A and M lineages has received the greatest attention with numerous studies using a wide array of markers ranging from morphology (Cornuet & Fresnaye 1989; Arias *et al.* 2006; Miguel *et al.* 2011), allozymes (Smith & Glenn 1995; Arias *et al.* 2006), mitochondrial DNA (Smith *et al.* 1991; Garnery *et al.* 1995, 1998a; Franck *et al.* 1998; De la Rúa *et al.* 2001, 2004, 2005; Arias *et al.* 2006; Miguel *et al.* 2007; Cánovas *et al.* 2008; Pinto *et al.* 2012, 2013), microsatellites (Franck *et al.* 1998; Garnery *et al.* 1998b; De la Rúa *et al.* 2002, 2003; Miguel *et al.* 2007, 2011; Cánovas *et al.* 2011) to SNPs (Chávez-Galarza *et al.* 2013).

Differential and complex diversity patterns emerged from the numerous biparental and maternal surveys of Iberian honey bees and the underlying processes shaping genetic variation remain controversial. Arguments based on selection, demography, and contemporary human-mediated processes have been favoured by different authors. Morphology (Ruttner *et al.* 1978; Cornuet & Fresnaye 1989) and the allozyme malate dehydrogenase (Nielsen *et al.* 1994; Smith & Glenn 1995; Arias *et al.* 2006) exhibited a smooth latitudinal cline extending from North Africa to France supporting a hypothesis of primary intergradation (Ruttner *et al.* 1978). In contrast, the abrupt transition from highly divergent M mitotypes in the northeastern half of Iberia to A mitotypes in the southwestern half (Garnery *et al.* 1995; Franck *et al.* 1998; Arias *et al.* 2006; Miguel *et al.* 2007; Cánovas *et al.* 2008) was more compatible with secondary contact (Smith *et al.* 1991). To complicate matters further, microsatellites did not capture the signal of a contact zone in Iberia (Franck *et al.* 1998; Garnery *et al.* 1998b; Cánovas *et al.* 2011; Miguel *et al.* 2011), but detected instead a sharp disruption between Iberian and northern African populations (Franck *et al.* 1998). This latter finding prompted a third hypothesis that explained occurrence of A mitotypes in Iberia by human-assisted introductions of African colonies during Muslim occupation, with selection acting to maintain

the morphological, allozymic and maternal clines (Franck *et al.* 1998). The hypothesis of selection as the driving force shaping the Iberian cline was recently addressed in a genome-wide SNP scan conducted in a fine-scale sample that covered the entire Iberian honey bee range (Chávez-Galarza *et al.* 2013). This study detected signatures of selection in the Iberian honey bee genome, suggesting that this evolutionary force has had an important role in structuring Iberian honey bee diversity.

Here, we built from those findings to provide, at both geographic and genomic levels, the most comprehensive characterization of the Iberian honey bee diversity patterns performed until now. We employed multiple clustering approaches and cline analysis to examine the genome-wide SNP data set using a population-genomics framework. After analysing the patterns of variation generated by the complete SNP data set, we removed any SNPs putatively associated with selection identified by Chávez-Galarza *et al.* (2013) and then used concurrently a mtDNA locus and 309 remaining neutrally tested SNPs to address the following questions: (i) How effective are SNPs in capturing clinal variation? – a signal that microsatellites have failed to detect, (ii) How concordant are the patterns generated by the complete and the neutral SNP data sets? (iii) Do neutral SNPs capture the clinal signal? (iv) Does the mtDNA marker confirm the presence of a cline formed by two highly divergent lineages, as documented by earlier studies? and How concordant are the patterns of neutral and maternal variation? If variation originated from secondary contact, then we expect neutral SNPs to detect clinal patterns and multiple coincident clines. In contrast, coincident neutral clines are not expected if variation originated via primary intergradation. Further evidence for secondary contact will come from comparisons of mtDNA and nuclear DNA. If a maternal cline formed by two highly divergent lineages is observed and this cline is paralleled by nuclear DNA variation, then there is strong support for secondary contact. However, given that Iberian honey bees are managed organisms, it is possible that human-mediated processes have obscured historical patterns. Therefore, we also asked the question: (v) To what extent do contemporary human-mediated forces influence the Iberian honey bee structure?

Methods

Sampling

Sampling in Iberia was conducted in 2010 across three north–south transects (Fig. 1). One transect extended along the Atlantic coast [Atlantic transect (AT)], one through the centre [central transect (CT)], and another along the Mediterranean coast [Mediterranean transect (MT)]. A total of 711 honey bee haploid males were collected in the three transects from 23 sites (AT=8, CT=9; MT=6) representing 237 apiaries and 711 colonies. Samples were stored in absolute ethanol at -20 °C until molecular analysis. Global positioning system (GPS) coordinates were recorded in the field for each apiary. Further details on sampling procedure can be found in Chávez-Galarza *et al.* (2013).

DNA extraction and marker genotyping

Using a phenol–chloroform–isoamyl alcohol (25:24:1) protocol (Sambrook *et al.* 1989), total DNA was extracted from the thorax of the 711 individuals, each representing a single colony. Analysis of mtDNA was based on the very popular $tRNA^{leu}$-cox2 intergenic locus. This region was amplified using the primers and PCR conditions detailed elsewhere (Garnery *et al.* 1993). After PCR amplification, the products were sequenced in both directions. Analysis of nuclear DNA was based on SNPs, which were genotyped using Illumina's BeadArray Technology and the Illumina GoldenGate® Assay with a custom Oligo Pool Assay (Illumina, San Diego, CA, USA) following manufacturer's protocols. Genotype calling was performed using ILLUMINA'S GENOMESTUDIO® data analysis software.

Of the 1536 SNPs included in the GoldenGate array, 383 passed the quality filtering and were polymorphic in *A. m. iberiensis*, using a cut-off criterion of > 0.98 for the most common allele. The 383 SNPs (referred hereafter as the complete SNP data set) were scanned for signatures of selection using four multiple-population F_{ST}-based methods and the spatial method matSAM (Chávez-Galarza *et al.* 2013). The 74 outlier loci detected by at least one of the five methods were removed from the complete data set, leaving 309 neutrally tested SNP loci (referred hereafter as the neutral SNP data set). Chromosomal locations in honey bee linkage groups and a summary of physical distances of the SNPs are shown in Table S1 (Supporting information). Details of SNP genotyping and detection of outlier loci can be found in Chávez-Galarza *et al.* (2013).

Mitochondrial DNA analysis

The sequences produced for the $tRNA^{leu}$-cox2 intergenic locus were aligned using MEGA version 5.03 (Tamura *et al.* 2011). For the identification of mitotypes, the sequences were compared with those available in GenBank. Each mitotype was then assigned to western European (M), eastern European (C), and African (A) lineages or to an African sublineage (A_I, A_{II}, A_{III}), according to the complex architecture of the $tRNA^{leu}$-cox2

Fig. 1 Map of the Iberian Peninsula showing the centroids of the sampling sites at each transect (AT, Atlantic; CT, central; MT, Mediterranean), sample size per site (n), site codes (CT1 to MT6), and the west–east transect (dashed line) traced from Lisbon to Girona for the geographic cline analysis (see Fig. 6). The approximate location of putative inferred refugia for the Iberian honey bee and other Iberian fauna that supports them (adapted from Gómez & Lunt 2007) is also shown.

intergenic region described elsewhere (Garnery *et al.* 1993; Franck *et al.* 1998, 2001; Rortais *et al.* 2011; Pinto *et al.* 2012). In short, the intergenic region is composed of two elements: the P (size varies between ~53 and 68 bp) and the Q (size varies between ~194 and 196 bp). A combination of point mutations and indels in the P element distinguishes honey bee subspecies from different lineages. The number of Q elements can vary between one and four, although the number of repeats is not lineage specific. Given the highly variable size of the sequences resulting predominantly from the variable number of Q elements, the sequences were trimmed at the end of the first Q with additional Qs coded as present/absent. Therefore, the ~627-bp sequence fragment analysed herein encompassed the 5′ end of the $tRNA^{leu}$ gene, the P element, the first Q element, the coding relative to the other Q elements, and the 3′ end of the cox2 gene. Relationships among the sequences were inferred using the median-joining network algorithm (Bandelt *et al.* 1999), as implemented in the program NETWORK version 4.6.1.1 (Fluxus Engineering, Clare, UK; http://www.fluxus-engineering.com). The trimmed ~627-bp sequences examined in this study did not diverge from those downloaded from GenBank. Therefore, GenBank sequences were not included in the network analysis.

Estimation of structure by nonspatial approaches

Structure was inferred from the complete and the neutral SNP data sets using two approaches: the Bayesian model-based STRUCTURE and the model-free discriminant analysis of principal components (DAPC). These approaches were used to estimate the proportion of an individual's genome (*Q*) that originated from a given genetic group or cluster.

The Bayesian clustering approach was implemented in STRUCTURE 3.4 (Pritchard *et al.* 2000) for haploid data using the admixture ancestry and correlated allele frequency models run with the unsupervised option. The program was set up for 750 000 Markov chain

Monte Carlo (MCMC) iterations after an initial burn-in of 250 000, which was sufficient to reach convergence. Over 20 independent runs for each number of clusters (K), from 1 to 7, were performed to confirm consistency across runs. The Greedy algorithm, implemented in the software CLUMPP 1.1.2 (Jakobsson & Rosenberg 2007), was used to compute the pairwise 'symmetric similarity coefficient' between pairs of runs and to align the 20 runs for each K. The means of the permuted results were plotted using the software DISTRUCT 1.1 (Rosenberg 2004). The optimal K value was determined using Evanno's ΔK (Evanno *et al.* 2005) and Campana's ΔF_{st} methods (Campana *et al.* 2011) in STRUCTURE HARVESTER web v0.6.93 (Earl & Von Holdt 2012) and CORRSIEVE 1.6-8 package (Campana *et al.* 2011), respectively.

The DAPC clustering approach was implemented in ADEGENET 1.3-9 package for R (Jombart 2008). Simulation studies have shown that DAPC performs as well or better than STRUCTURE, particularly under more complex structuring scenarios (Jombart *et al.* 2010; Klaassen *et al.* 2012). DAPC provides a description of the genetic clustering using coefficients of the alleles (loadings) in linear combinations and seeks to maximize between-groups variance and minimize within-group variance in these loadings (Jombart *et al.* 2010). Successive K-means clustering runs (from 1 to 40) were also incorporated in the analysis to estimate the optimal number of distinct clusters (K) based on the Bayesian information criterion (BIC). The optimal K value is associated with the lowest BIC value (Jombart *et al.* 2010).

Estimation of structure by spatial approaches

Spatial structure was inferred from the complete and the neutral SNP data sets using two approaches that explicitly incorporate information on geographic coordinates for genotyped individuals: the model-free multivariate spatial principal component analysis (sPCA) and the Bayesian model-based TESS. The sPCA is a modification of PCA which takes into account not only the genetic variance of individuals or populations but also their spatial autocorrelation (measured by Moran's I). This approach disentangles global structures (clines, patches or intermediates) from local structures (strong genetic differences among neighbours), and from random noise (random distribution of allelic frequencies among individuals or populations on a connection network). While global structures display positive spatial autocorrelation (high positive eigenvalue), local structures display negative spatial autocorrelation (high negative eigenvalue) (Jombart *et al.* 2008). The sPCA was performed in ADEGENET using the K-nearest neighbours to model the spatial connectivity among individuals. To test for global and local structures, a Monte Carlo test was implemented using 10 000 permutations.

The Bayesian model-based clustering approach implemented by the software TESS (Chen *et al.* 2007) incorporates spatial population models assuming geographic continuity of allele frequencies by including the interaction parameter Ψ, which defines the intensity of two neighbouring individuals belonging to the same genetic cluster. The incorporation of trend surface and a Gaussian autoregressive residual term allows for capturing global and local patterns. The software TESS 2.3.1 was run for haploid data using the convolution admixture model (BYM), correlated allele frequency and a trend degree surface of 1. A Euclidean distance matrix was used to weight the spatial connectivity among individuals. Five runs were carried out at each K, from 2 to 7, with 5 000 000 MCMC total sweeps including a burn-in of 1 250 000 sweeps. For each run, the deviance information criterion (DIC) was calculated and the values of all runs were averaged and plotted against K. The first lower DIC value represents the optimal K for the data. As in STRUCTURE, the software programs CLUMPP and DISTRUCT were used to obtain the average matrix of membership proportions (Q) for each K and for graphical representation.

Geographic cline analysis

Geographic clines were estimated for mtDNA frequency, individual SNPs frequency, and mean Q per sampling site inferred from both the complete and the neutral SNP data sets by STRUCTURE at $K = 2$. Cline analysis of individual SNPs was performed on a subset of loci with an absolute allele frequency difference > 0.2. The rationale behind this choice was that the SNP panel used in this study was ascertained from the reference genome of *A. mellifera* (sequenced from the North American DH4 strain, which was primarily *A. m. ligustica*, a subspecies belonging to the C lineage) and genome sequence traces of Africanized honey bees (largely the African *A. m. scutellata* admixed with the genomes of *A. m. ligustica* and the M-lineage *A. m. mellifera*), and is thus underrepresented for diagnostic SNPs and SNPs with large allele frequency differences between the two maternal lineages identified in Iberian honey bees. Using a subset of loci with larger allele frequency difference between the groups, we expected to increase the ancestry information of individual loci for cline analysis. Of the 383 SNPs, 33 (17 neutral and 16 under selection, as identified by Chávez-Galarza *et al.* 2013) conformed to the frequency criterion.

Sampling sites were arranged along a transect beginning at the westernmost location (Lisbon, Portugal) and ending at the easternmost location (Girona, Spain)

(Fig. 1). Each sampling site was assigned a distance along this transect, which corresponded to the shortest straight-line distance between it and Lisbon, calculated using the 'harversine' approach (www.movable-type.co.uk/scripts/latlong.html). The cline shape was modelled using the package HZAR v2.5 (Derryberry *et al.* 2014), which fits allele frequency data to equilibrium geographic cline models (Szymura & Barton 1986, 1991; Barton & Gale 1993; Gay *et al.* 2008) using the Metropolis–Hastings Markov chain Monte Carlo algorithm. The following cline shape parameters were estimated: centre (c, distance from sampling location), width (w, 1/maximum slope), delta (δ, distance between the centre and the tail) and tau (τ, slope of the tail). The allele frequencies at the top and bottom of the cline (*Pmin* and *Pmax*) were either fixed or free to vary. Three sets of five cline models were fitted: model set 1 had no scaling (*Pmin* = 0, *Pmax* = 1), model set 2 had fixed scaling (*Pmin* = observed minimum, *Pmax* = observed maximum), and model set 3 allowed *Pmin* and *Pmax* to vary. Within each model set, scaling and tails were fixed or free to vary. These models were compared to a null model of no clinal transition using the Akaike Information Criterion corrected (AICc). The best-fitting model had the lowest AICc value. To evaluate coincidence among cline centre positions and concordance among cline widths, the composite likelihood method (Phillips *et al.* 2004) was used. Likelihood profiles were constructed for both c and w to compare alternative hypotheses across loci: H1, all loci are characterized by statistically indistinguishable c and w values and are likely to share a common c/w; and H2, each locus has its own independent c and w values. Composite log-likelihood profiles were constructed by summing log-likelihood (ML) profiles for all individual SNP loci ML(H1). This composite log-likelihood profile was compared to the sum of all maximum-likelihood estimates for individual SNP loci ML(H2) using a likelihood ratio test (LRT). If the clines of individual SNP loci coincide and have the same c/w values, ML(H1) is not significantly different from ML (H2) (ML = ML(H2) – ML(H1) $\approx$ 0). Conversely, if the clines do not coincide, ML(H1) is significantly smaller than ML(H2) (ML > 0). The significance of any difference of ML(H1) and ML(H2) was determined using a chi-square test with $n-1$ degrees of freedom ($\alpha = 0.05$). This approach was similarly employed to evaluate coincidence and concordance of mtDNA and both SNP data sets.

Linkage disequilibrium and genetic diversity

Linkage disequilibrium (LD) between all pairs of neutral SNPs was estimated using the statistic r^2 (Hill & Robertson 1968), as implemented by the software DNASP 5.10.1 (Librado & Rozas 2009). Significant LD was identified at the 5% level using Fisher's exact test. Unbiased haploid genetic diversity (*uh*) for the neutral SNPs was calculated using the program GENALEX 6.5 (Peakall & Smouse 2012). Values of LD and *uh* were calculated for each sampling site and then projected along the Lisbon–Girona transect (Fig. 1).

Statistical tests

Differences in individuals' Qs between clustering approaches were assessed using the Mann–Whitney–Wilcoxon test (Wilcoxon 1945; Mann & Whitney 1947). Genetic structure inferred by the different clustering approaches was compared using Pearson's correlation coefficient (r). Whenever applicable, statistical significance levels were adjusted for multiple comparisons using the Bonferroni procedure to correct for type I error (Weir 1996). These analyses were implemented in R (R Development Core Team 2013).

Results

Structure estimated by nonspatial approaches

Genetic structure inferred from the 711 Iberian honey bee individuals and the 309 neutral SNP data set using the Bayesian model-based clustering algorithm, implemented in STRUCTURE, and the model-free DAPC clustering algorithm, implemented in ADEGENET, is shown in Fig. 2a (for $K = 2$) and Fig. S1 (Supporting information) ($K = 3$ to 5). The optimal number of clusters (K) varied between two, when estimated by ΔF_{st} and BIC, and four, when estimated by ΔK (Fig. S2, Supporting information). Incongruent optimal K values are often obtained by different methods (Campana *et al.* 2011), especially in the presence of low levels of population differentiation (Waples & Gaggiotti 2006), which is the case of the Iberian honey bee (global $\Phi_{PT} = 0.020$; pairwise Φ_{PT} values ranged from 0.000 to 0.046, but see Table S2 and Fig. S3, Supporting information for pairwise comparisons across the study area). Given that two of three measures agreed on an optimal $K = 2$ and the presence of two maternal lineages in Iberia (this and previous studies), it is likely that the number of clusters that best represents the maximum population structure is two.

At the optimal $K = 2$ (Fig. 2a, Fig. S4, Supporting information), a concordant geographic pattern was produced by DAPC and STRUCTURE ($r = 0.79$, P-value = 0.0000 for individual Q values), although a deeper subdivision was inferred by the former than by the latter clustering approach, as measured by Q (P-value = 0.0000 for comparisons of individual Q values, Mann–Whitney–Wilcoxon test). Membership proportions

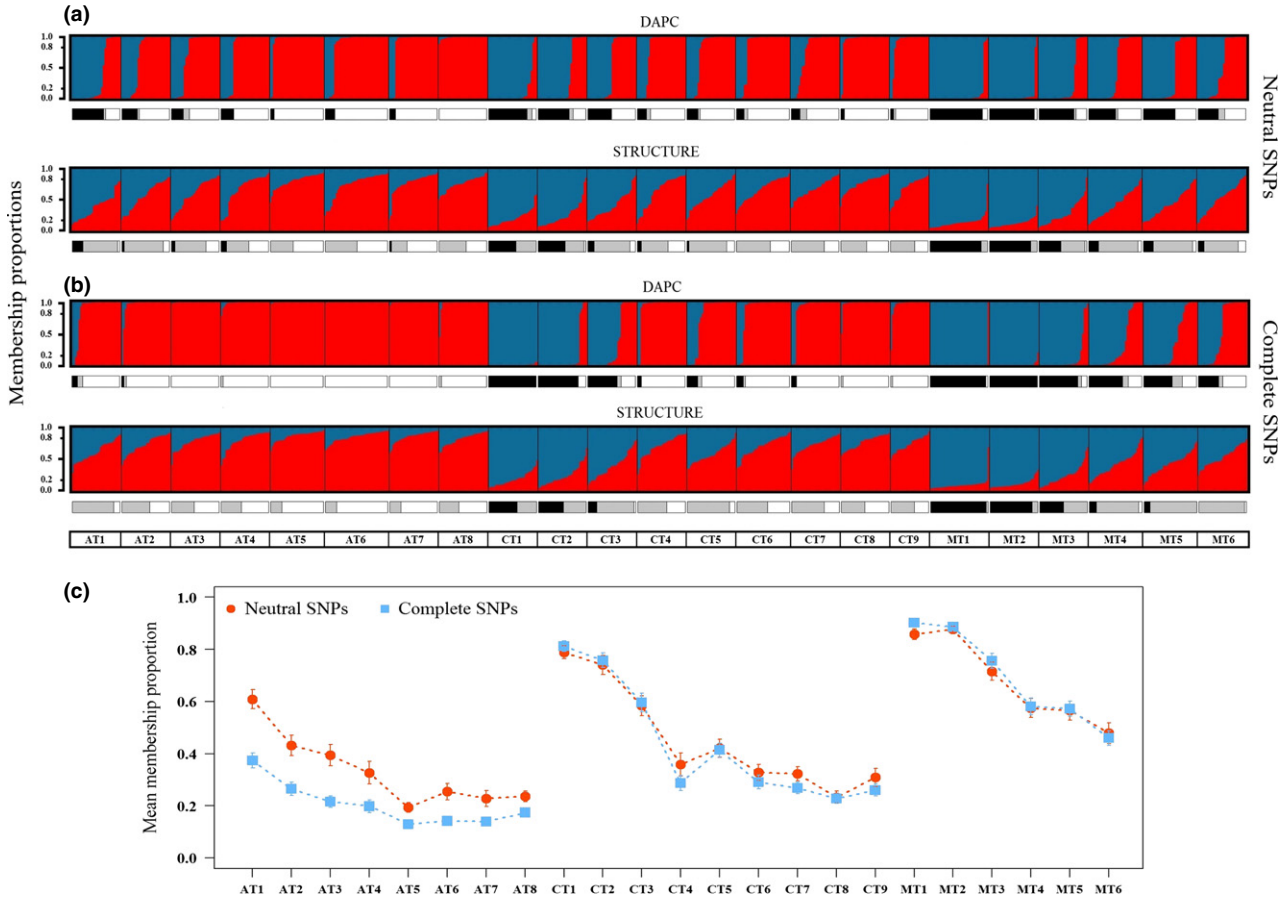


Fig. 2 Population structure of *Apis mellifera iberiensis* estimated by DAPC and STRUCTURE at $K = 2$ clusters. The 23 sampling sites are arranged from north (AT1, CT1, MT1) to south (AT8, CT9, MT6) in each of the three transects (AT, Atlantic; CT, central; MT, Mediterranean). Plots represent each of the 711 individuals by a vertical bar partitioned into two coloured segments (blue and red) corresponding to membership proportions (Q) in each of the two clusters. Black lines separate individuals from the 23 sampling sites, which are arranged from high Q (left) to low Q (right) in the blue cluster. The frequency of individuals per sampling site exhibiting $Q \geq 0.80$ (black), $0.20 > Q > 0.80$ (grey), and $Q \leq 0.20$ (white) in the blue cluster is indicated below each plot. Structure estimated from (a) the neutral SNP data set (309 loci) and (b) the complete SNP data set (309 neutral plus 74 putatively selected loci identified by Chávez-Galarza *et al.* 2013). (c) Mean membership proportion (±SE) in the blue cluster inferred from the neutral and complete SNP data sets with STRUCTURE for each sampling site.

estimated by STRUCTURE showed that most individuals (ranging from 53.3% in CT2 to 88.9% in MT1; see the bar below each clustering plot in Fig. 2a) from sampling sites near the Pyrenees were assigned with high posterior probability ($Q \geq 0.80$) to the blue cluster. The percentages increased considerably when Q was inferred by DAPC, ranging from 50.0% in CT3 to 93.3% in MT2. Individuals with $Q \geq 0.80$ in the red cluster were common in the southern sampling sites of the Atlantic transect (ranging from 60.6% in AT5 to 63.3% in AT7) and rare in the Mediterranean transect (ranging from 0% in MT1-2 to 13.3% in MT6). However, again, the percentages increased considerably when Q was inferred by DAPC. Individuals exhibiting admixed proportions (Q to any cluster ≤ 0.80) prevailed in the northern part of the Atlantic transect and in the southern part of the central and Mediterranean transects when inferred by STRUCTURE, although they were rare when inferred by DAPC (Fig. 2a).

When genetic structure was inferred from the complete SNP data set (Fig. 2b), a deeper phylogeographical signal was captured by both clustering approaches, as measured by Q (P-value = 0.0000 for comparisons of individual Q values inferred from the complete and neutral data sets with STRUCTURE and DAPC; Mann–Whitney–Wilcoxon test). Nonetheless, inclusion of the 74 putatively selected SNPs in the data set did not qualitatively change the overall geographic patterns of hybridization across Iberia, while providing additional support for an optimal $K = 2$, this time simultaneously estimated by the three methods ΔK, ΔF_{ST} and BIC (Fig. S2, Supporting information).

The mean *Q* estimated with STRUCTURE from both the neutral and complete SNP data sets is shown at the sampling site level in Fig. 2c (see Fig. S5, Supporting information for the corresponding DAPC plot). While further confirming the nearly concordant patterns across Iberia, this representation revealed a more abrupt transition from the blue to the red cluster in the central and Mediterranean transects than in the Atlantic transect, which suggests a contact zone located towards the northeastern part of Iberia.

Comparing maternal pattern with neutral structure

A median-joining network of a ~627-bp fragment of the tRNA^{leu}-cox2 intergenic mitochondrial region confirms the presence of the two highly divergent African (A) and western European (M) lineages in Iberia (Fig. 3a). The two lineages show a highly structured geographic pattern of distribution in Iberia. Mitotypes belonging to the M lineage were predominant in the northeastern half, whereas mitotypes belonging to the A lineage were fixed or almost fixed in the southwestern half of Iberia. While some sampling sites displayed a mixture of M and A mitotypes, the geographic distribution of the maternal lineages reveals a sharp northeastern–southwestern trend.

Membership proportions inferred by STRUCTURE (Fig. 3b, c) and DAPC (Fig. S6, Supporting information) from neutral SNPs were contrasted with mtDNA data, at both sampling site and individual levels. At the

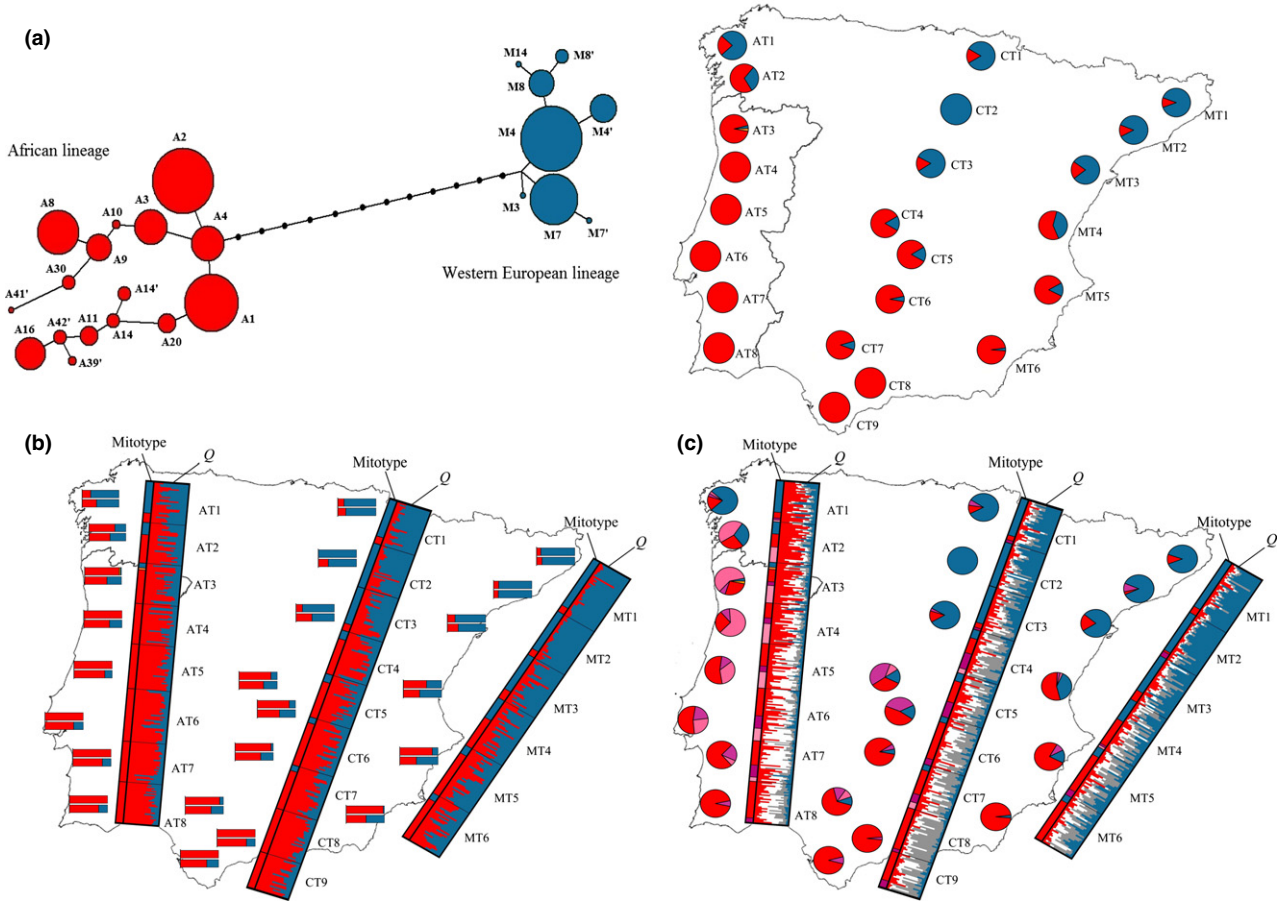


Fig. 3 Maternal pattern and estimated structure inferred from the neutral SNP data set (309 loci) with STRUCTURE. Patterns of variation are displayed at both individual and sampling site level for each transect (AT, Atlantic; CT, central; MT, Mediterranean). (a) Median-joining network relating the mtDNA sequences of a ~627-bp fragment of the tRNA^{leu}-cox2 intergenic region. The African (A) and western European (M) mitotypes form two divergent lineages. The sizes of the circles size are proportional to the mitotype frequencies. The pie charts displayed in the Iberian map at the right side show the frequencies of the A and M mitotypes at each sampling site. (b) Patterns shown at $K = 2$ clusters. Vertical plots display the mitotype (A in red; M in blue; C in orange, one single individual in AT3) and the membership proportions (*Q*) for each of the 711 individuals. Horizontal bar plots show mitotype frequencies (top) and the mean *Q* in blue and red clusters (bottom) at each sampling site. (c) Patterns shown at $K = 4$ clusters. Vertical plots display the mitotype and *Q* for each of the 711 individuals. Maternal data are represented by M lineage (blue), C lineage (orange) and A sub-lineages (A_I in red, A_{II} in magenta, and A_{III} in pink). Pie charts show mitotype frequencies at each sampling site.

sampling site level, the partitioning of neutral SNP variation into two clusters corresponded remarkably to M and A maternal lineages ($r = 0.81$ for both DAPC and STRUCTURE *vs* mtDNA; P-value < 0.0000). At the individual level, the correlations were weaker ($r = 0.46$ for DAPC, $r = 0.60$ for STRUCTURE), yet significant (P-value < 0.0000), suggesting differential gene flow among genomic compartments.

Genome partitioning of individuals at greater values of K produced increasingly complex patterns (Fig. 3c and Fig. S1, Supporting information). At $K = 4$, a pronounced east–west structuring of neutral variation was revealed. The Atlantic populations were clearly distinct from the other populations, and a north–south trend becomes more apparent in this transect. The nuclear pattern is consistent with maternal variation partitioned into African sublineages (Fig. 3c). Sublineage A_{III} mitotypes were common in northern Atlantic populations and were gradually replaced by sublineage A_I mitotypes towards the south. In contrast to Atlantic populations, sublineage A_{III} mitotypes were virtually absent in populations of central and Mediterranean transects, which were dominated by sublineage A_I mitotypes.

Structure estimated by spatial approaches

A number of studies have questioned the use of STRUCTURE for studying populations exhibiting continuous spatial distribution of genetic diversity (Serre & Pääbo 2004; Rosenberg *et al*. 2005). To address this issue, patterns of variation in Iberia were further investigated using spatially explicit approaches implemented by sPCA (Fig. 4) and TESS (Fig. 5).

Analysis of neutral SNPs using sPCA showed that one global axis and one local axis were retained, indicating the existence of both global and local spatial structures in Iberia. The interpolation of the first global score, which was associated with a strong autocorrelation (Moran's $I = 0.639$), detected two clusters forming a cline (Fig. 4a) concordant with nonspatial approaches. The second global score (Moran's $I = 0.560$) clearly differentiated the four northernmost sampling sites of the Atlantic transect and the southern half of central and Mediterranean transects (Fig. 4b). The third global score (Moran's $I = 0.443$) further partitioned the Atlantic populations into two groups (north and south) and the southern half of central transect (Fig. 4c). The northern half of the central transect was differentiated by the

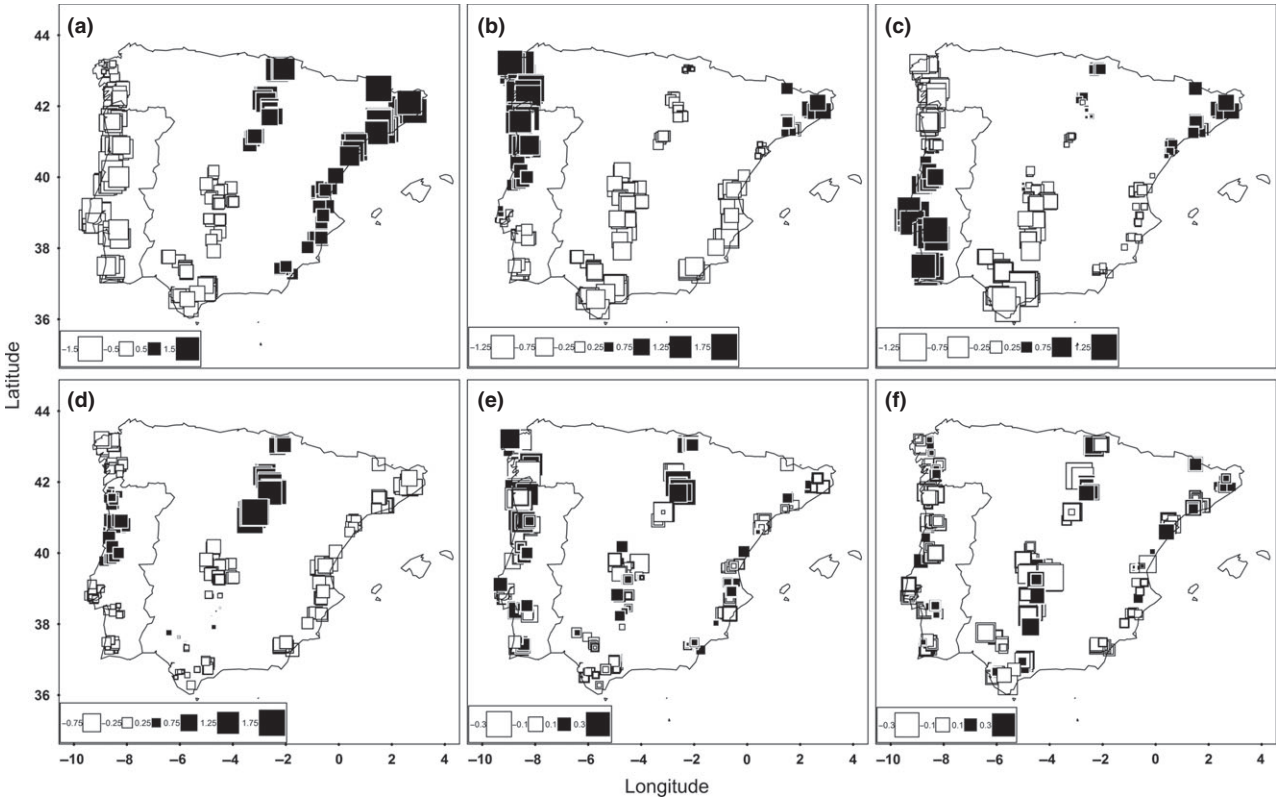


Fig. 4 Analysis of global and local structures, among 711 individuals of *A. m. iberiensis* from 23 sampling sites, by spatial principal component analysis (sPCA) using 309 neutral SNPs. Each square represents the score of an individual, which is positioned by its spatial coordinates. (a–d) The first four global scores of sPCA. (e–f) The first two local scores of sPCA.

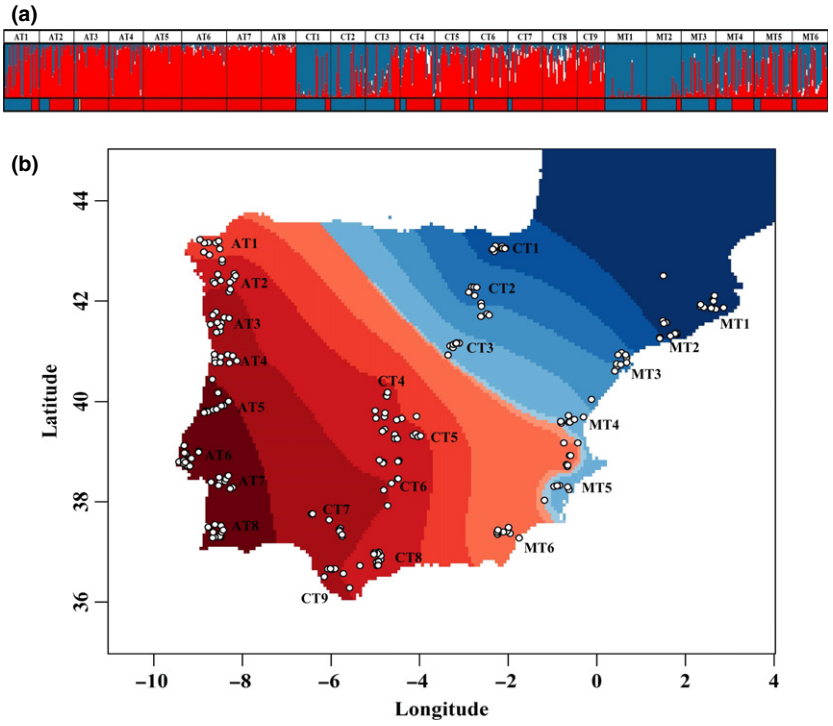


Fig. 5 Spatially explicit analysis implemented by the software TESS for the 711 individuals of *Apis mellifera iberiensis* using the neutral SNP data set (309 loci). (a) Plot of individuals' *Q* at the optimal *K* = 3 clusters. Each of the 711 individuals included in the analysis is represented by a vertical bar partitioned into three coloured segments (blue, red and white) corresponding to *Q* in each of the three clusters. Maternal data (M lineage in blue, C lineage in orange, and A lineage in red) are shown at the bottom. Sampling sites and individuals within sampling sites are arranged as in Fig. 3. (b) Map of the Iberian Peninsula showing the two major clusters ($Q \geq 0.5$) interpolated by TESS. Dots represent the locations of sampled apiaries across the Atlantic (AT1-8), central (CT1-9) and Mediterranean (MT1-6) transects.

fourth global score producing a Moran's $I = 0.392$ (Fig. 4d).

While the global test corroborated the presence of global spatial structure (max(t) = 0.0017; *P*-value = 0.0001), there was also structure at the local level (max (t) = 0.0019; *P*-value = 0.0001). The first local score (Moran's $I = -0.075$) highlighted the differences among individuals of northern Atlantic and central transects (Fig. 4e), while the second local score (Moran's $I = -0.071$) differentiated the individuals from sampling sites in the middle part of the central transect (Fig. 4f).

The additional spatial approach performed using TESS further confirmed the neutral patterns obtained previously (Fig. 5). Two major clusters that largely overlapped those of nonspatial approaches ($r = 0.76$ for STRUCTURE *vs* TESS and $r = 0.64$ for DAPC *vs* TESS using individual *Q* values, *P*-value < 0.0000) were identified by TESS at each simulated *K* (Fig. S7, Supporting information). At the optimal *K* = 3 (Fig. S8, Supporting information), one additional minor cluster (mean *Q* = 0.023 in the white cluster) further partitioned the nuclear genomes of individuals mainly from the southern portion of the central transect (Fig. 5a). While TESS supported the major northeastern–southwestern cline and the contrasting patterns exhibited by Atlantic and Mediterranean populations, it did not capture the partitioning within the Atlantic transect, which was detected by the other clustering approaches and by mtDNA analysis.

The spatial patterns inferred from the complete SNP data set using both sPCA and TESS were largely concordant with those inferred from the neutral SNP data set, although, as observed with the nonspatial clustering approaches, a deeper phylogeographical signal was captured by the complete SNP data set (Figs S9 and S10, Supporting information).

Geographic cline analysis

The geographic clines were modelled for 33 (17 neutral and 16 selected) individual SNPs, mean *Q* obtained with the complete and neutral SNP data sets, and mtDNA (Fig. 6, Fig. S11, Supporting information). There was considerable variation in the identity of the best-fitting model among individual SNPs, SNP data sets and mtDNA (Table S3, Supporting information).

The model 'Pmin/Pmax observed – no tails' was fitted to 16 of the 33 SNPs and to the neutral SNP data set, whereas 'Pmin/Pmax fixed – right tail' and 'Pmin/Pmax fixed – no tails' were fitted to the mtDNA and the complete SNP data set, respectively. The AICc values obtained for the null model of no clinal variation were higher for mtDNA (cline model = 255.1, null model = 516.3), complete SNP data set (cline model = 25.2, null model = 203.5) and neutral SNP data set (cline model = 33.2, null model = 137.2) than for any individual SNP (Table S3, Supporting information).

Estimates of cline centre positions and widths for the 33 individual SNPs, complete SNP data set, neutral SNP data set and mtDNA were highly variable (Table S3, Supporting information). Coincidence analysis of the 33 SNPs revealed that 18 (Table S3, Supporting information), of which nine were neutral, could be constrained to share a common centre ($LRT_{same\text{-}diff.}$ = 26.88, 17 d.f., *P*-value > 0.05) at the consensus position of 665.2 km, as estimated by the likelihood profiles. The consensus centre of the 18 SNPs was coincident with those estimated for mtDNA (706.7 km) and for the complete (714.7 km) and neutral (725.7 km) SNP data sets ($LRT_{same\text{-}diff.}$ = 0–3.03, 1 d.f., *P*-value > 0.05 for all pairwise comparisons). Concordance analysis of the 33 SNPs revealed that 23 (Table S3, Supporting information), of which 11 were neutral, exhibited a similar width ($LRT_{same\text{-}diff.}$ = 1.60, 22 d.f., *P*-value > 0.05) of 1350 km. The consensus width of the 23 SNPs was concordant ($LRT_{same\text{-}diff.}$ = 0.3 – 3.18, 1 d.f., *P*-value > 0.05 for all pairwise comparisons) with those estimated for the complete (1283.5 km) and neutral SNP data sets (1047.6 km), but not with that estimated for the mtDNA (580.9 km), which was significantly narrower ($LRT_{same\text{-}diff.}$ = 10.10–26.56, 1 d.f., *P*-value < 0.05 for all pairwise comparisons).

Linkage disequilibrium and genetic diversity

Genome-wide analysis of linkage disequilibrium (LD) between all possible pairs of neutral SNPs in each sampling site produced low levels of LD with mean r^2 values varying between 0.014 and 0.045 (Table S4, Supporting information). From a total of 701,362 pairwise comparisons, 14 687 pairs (2.09%), ranging from 1.49% to 3.41%, exhibited significant LD before Bonferroni correction (a single pair in CT4 remained significant after Bonferroni correction). Pairwise comparisons performed by linkage group also produced low LD values (data not shown). Levels of unbiased haploid diversity (*uh*) were low, ranging from 0.281 in the Atlantic transect (AT8) to 0.313 in the Mediterranean transect (M6, Table S4, Supporting information). Interestingly, despite the low levels of LD and *uh* across the study area, a trend of elevated values was observed towards the centre of the cline and overlapping the consensus centre location (Fig. 7).

Discussion

Genetic studies of Iberian honey bees have revealed complex and often incongruent patterns of variation, which have led to competing hypothesis of primary intergradation (Ruttner *et al.* 1978) and secondary contact (Smith *et al.* 1991) as the leading mechanisms shaping patterns of variation. We examined a large number of individuals with a maternal locus and genome-wide SNPs and provided the most comprehensive portrait of clinal change across the entire Iberian honey bee distributional range. Our results support a signature of origin via secondary contact, which was still detectable despite intense beekeeping practices involving selective breeding and large-scale movement of colonies.

Nuclear and maternal patterns and cline origin

The multiple clustering approaches and the geographic cline analysis implemented on genome-wide SNPs collectively revealed a well-defined clinal pattern bisecting Iberia along a northeastern–southwestern axis, contrasting with the lack of microsatellite structure documented earlier (Franck *et al.* 1998; Cánovas *et al.* 2011; Miguel *et al.* 2011).

The most commonly suggested mechanisms underlying clinal patterns in gene frequencies are random genetic drift with isolation-by-distance effects, selection across an environmental gradient (primary intergradation), and secondary contact between previously isolated and genetically divergent populations. The SNP patterns exhibited by the Iberian honey bees could be explained by any of these mechanisms. However, several aspects of our data are more consistent with an alternative model of secondary contact and introgression between divergent populations previously isolated in glacial refugia, as proposed for a growing list of other Iberian taxa (reviewed by Gómez & Lunt 2007; Godinho *et al.* 2008; Gonçalves *et al.* 2009; Pinho *et al.* 2009; Miraldo *et al.* 2011; Carneiro *et al.* 2013; among others). First, while a deeper structure was retrieved by the complete SNP data set, excluding the 74 SNPs with signatures of selection (Chávez-Galarza *et al.* 2013) did not qualitatively change the clinal pattern of variation. A similar historical signal emerged when selected loci were removed from the genome-wide SNP data set.

Second, the geographic cline analysis revealed that a large proportion (9 of 17) of the neutral individual SNPs and both SNP data sets share a common cline

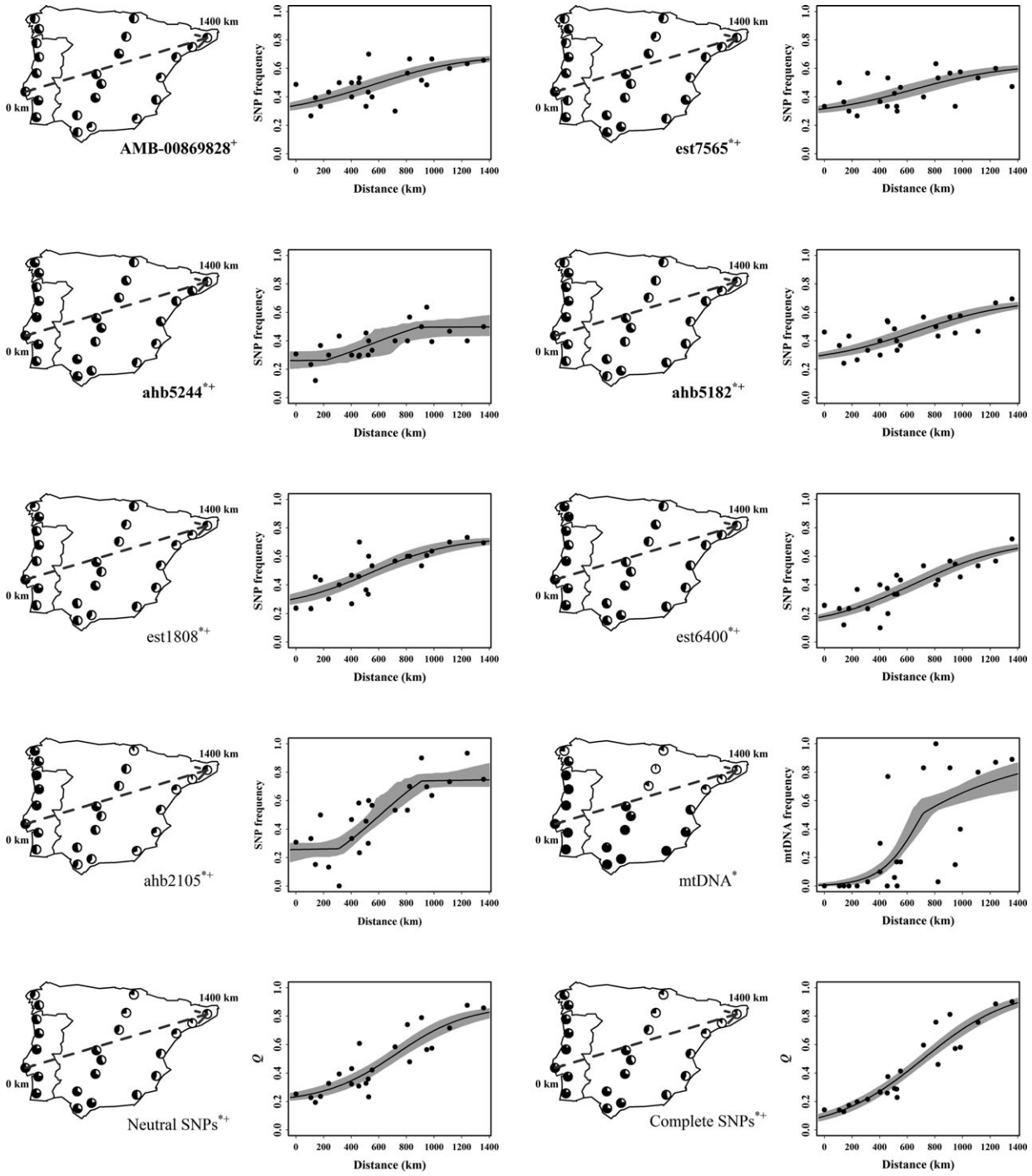


Fig. 6 Map of the Iberian Peninsula with pie charts summarizing frequency data for each sampling site and plot of maximum-likelihood geographic cline for four neutral SNP loci (marked in bold), three selected SNP loci (see Fig. S11, Supporting information for the remaining 26 SNPs), mtDNA, and the *Q* values estimated with STRUCTURE from the neutral and the complete SNP data sets. The symbols * and + indicate the loci or data sets with coincident centre and concordant width, respectively (see Table S3, Supporting information). The dashed line placed in each map represents the transect traced from Lisbon (0 km) to Girona (1400 km) for the geographic cline analysis.

centre, indicating considerable genome-wide coincidence. Existence of multiple coincident clines argues for secondary contact (Barton & Hewitt 1981), especially if some of the clines reflect changes in selectively neutral loci (Durrett *et al.* 2000). Simulations on the origin of contact zones show that a signature of secondary

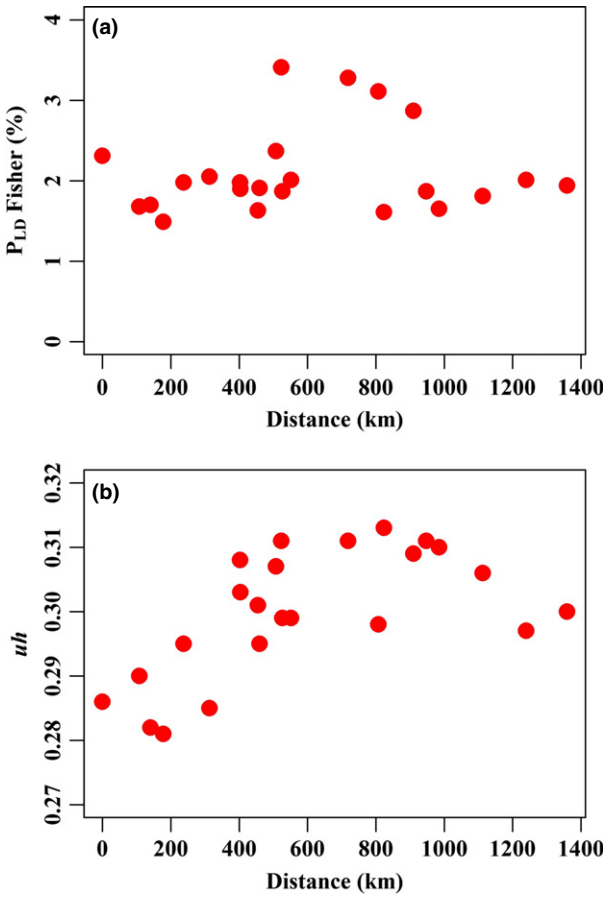


Fig. 7 (a) Percentage of pairs of neutral SNP loci showing significant linkage disequilibrium (LD) with Fisher's exact test (P_{LD} Fisher), before Bonferroni correction, and (b) unbiased haploid genetic diversity (*uh*) estimated from the neutral SNPs, both projected along the geographic cline.

contact, which is characterized by clinal variation at neutral loci and extensive disequilibrium at the centre of the contact zone, can persist for thousands of generations if neutral loci were tightly linked to selected loci (Durrett *et al.* 2000). If neutral loci were not tightly linked to selected loci, the initially steep clines, formed at the moment the two divergent groups meet, will gradually widen as the intermixing proceeds. In contrast, in the presence of primary intergradation, neutral loci will not vary clinally and disequilibrium between a neutral locus and a closely linked locus under selection will decay quickly (Durrett *et al.* 2000).

Further support for secondary contact is provided by LD and diversity patterns. Both parameters show a trend of elevated values towards the centre of the cline and overlapping the consensus centre location, as expected when two divergent populations meet. It should be noted, however, that the LD levels at the centre of the contact zone were unexpectedly low for recent contact, which may simply reflect the sparse distribution of the SNPs and the short scale of LD in honey bees. Indeed, the exceptionally high recombination rate in honey bees (Beye *et al.* 2006) would lead to a rapid decay of LD after admixture, as observed in the Africanization process in the New World (Pinto *et al.* 2005).

Finally, while the observed patterns can also be explained by alternative processes involving isolation by distance with some migration or by divergence in parapatry, the ultimate support for secondary contact comes from mtDNA. Congruent with previous surveys (Smith *et al.* 1991; Garnery *et al.* 1992, 1995; Franck *et al.* 1998; Cánovas *et al.* 2008; Miguel *et al.* 2007), two highly divergent mtDNA lineages (African and western European) were identified and these lineages form a cline that closely parallels that of genome-wide SNPs, as revealed by both the clustering and the cline analyses. The mtDNA cline centre coincided with that of both the complete and the neutral SNP data sets and with most individual SNPs (18 of 33). In contrast, the mtDNA cline width was not concordant with any of the SNPs (either individual or data sets), which largely formed wider clines. Narrower maternal clines could arise from stronger drift on the smaller N_e of the haploid marker (Polechova & Barton 2011), from divergent selection on mtDNA types (Yuri *et al.* 2009), or from the greater gene flow expected of nuclear loci (Endler 1977). Our results of narrower maternal clines compared to nuclear clines add to a body of research showing discordant cytonuclear transmission across contact zones (Gowen *et al.* 2014 and many references therein).

Whether secondary contact resulted from ancient range expansions from North Africa following climate amelioration of the last postglacial period (De la Rúa *et al.* 2002; Pinto *et al.* 2013) or from recent introductions of the North African subspecies *A. m. intermissa* by the Arabs during Muslim occupation (Franck *et al.* 1998) is a matter of debate. A STRUCTURE analysis found no signs of *A. m. intermissa* genes, belonging to the African lineage, in Iberian populations, excepting for a residual component detected in sampling sites CT8 and CT9 nearby the Strait of Gibraltar (see Fig. S3, Supporting information in Chávez-Galarza *et al.* 2013). This observation together with a report of deep differentiation between *A. m. iberiensis* and *A. m. intermissa* SNPs (Whitfield *et al.* 2006) means that a recent colonization event would have to be accompanied by a complete replacement of the nuclear, but not the mitochondrial, genomes of colonizers. This hypothesis assumes long-term male-biased gene flow, which would erode a signal of subdivision at the nuclear but not at the maternal level. The problem is that the latter scenario is not consistent with honey bee reproductive biology, as females have an important role in long-distance dispersal (Winston 1987). A much earlier event is

therefore more likely to have been responsible for the patterns we see today.

The complexities and incongruences of Iberian honey bee patterns revealed by distinct genetic markers suggest an ancient history of allopatric divergence in Iberian refugia followed by postglacial range expansions and secondary contact. Iberia served as an important refuge during the cold periods of the Pleistocene in Europe (reviewed by Weiss & Ferrand 2007). During this epoch, repeated cycles of contraction into and expansion out of multiple refugia shaped diversity patterns of great complexity in a variety of Iberian animal taxa (see Miraldo *et al.* 2011 and references therein), among which the honey bee is seemingly no exception. Estimates of genetic divergence from mtDNA (Arias & Sheppard 1996) and whole-genome nuclear DNA (Wallberg *et al.* 2014) suggest that the split among the four honey bee evolutionary lineages occurred between 670 000 and 300 000 years ago, respectively. Accordingly, colonization of Iberia across the Strait of Gibraltar (Ruttner *et al.* 1978; Whitfield *et al.* 2006; Han *et al.* 2012), from an origin in either Africa (Whitfield *et al.* 2006) or western Asia (Wallberg *et al.* 2014), likely occurred during Middle Pleistocene. Given dispersal abilities of the honey bees, it is plausible that they dispersed across the Iberian territory during interglacial periods and retreated to refugia during the glacial periods. Evidence from comparative phylogeography suggests that multiple refugia existed in Iberia ('refugia within refugia' paradigm of Gómez & Lunt 2007). Two such refugia, one in the Mediterranean coast of northeastern Spain, possibly close to the Ebro valley, and another in the Betic ranges of southern Spain, were inferred from overlapping subdivision patterns exhibited by several Iberian taxa (Gómez & Lunt 2007; see Fig. 1). The blue and the red clusters identified in Fig. 3 are consistent with the existence of these two putative refugia. Although the presence of multiple honey bee refugia is a tentative result, it is an idea that can be further explored and tested using the power of multiple gene genealogies analysis.

Influence of human-mediated processes in shaping variation

Complicating the interpretation of diversity patterns in honey bees are contemporary human-mediated processes. Honey bees native to Europe have long been subjected to human manipulation (Crane 1999), with a variable impact in their genetic composition (reviewed by De la Rúa *et al.* 2009). In western Europe north of the Pyrenees, human-mediated movements of colonies between lineages (introduction of commercial queens) promoted variable levels of C-lineage introgression (Jensen *et al.* 2005; De la Rúa *et al.* 2009; Soland-Reckeweg *et al.* 2009; Oleksa *et al.* 2011; Pinto *et al.* 2014) and even replacement of the native *A. m. mellifera* subspecies in some areas (Jensen *et al.* 2005). In contrast, the Iberian honey bee is relatively free of C-lineage genes (Miguel *et al.* 2007, 2011; Cánovas *et al.* 2008, 2011; Pinto *et al.* 2013). A single colony harbouring a C-derived mitotype was scored in this study, and no signs of introgression were detected at the nuclear level in that colony and in the remaining 710 (but see Figs S3 and S4, Supporting information in Chávez-Galarza *et al.* 2013).

While movements of colonies between lineages have not yet seriously threatened the Iberian honey bee genetic integrity, the lack of microsatellite structure (Franck *et al.* 1998; Cánovas *et al.* 2011; Miguel *et al.* 2011) has been interpreted as an indication of high levels of gene flow aided by within-lineage movements associated with transhumance (Cánovas *et al.* 2011). This interpretation, however, is inconsistent with our results that show congruent cytonuclear subdivision, local structure detected by the sPCA, and relatively low levels of LD, none of which support the large-scale influence of transhumance in the Iberian honey bee gene pool. A possible explanation for microsatellite patterns is that saturation of the mutation spectrum homogenized allele size distributions (Nauta & Weissing 1996). Such homogenization has been suggested before to explain a similar pattern in the European rabbit (Queney *et al.* 2001). Almost identical to what is observed for the Iberian honey bee, the European rabbit exhibits a northeastern–southwestern cline for mtDNA (Branco *et al.* 2000), allozymes (Campos *et al.* 2007; Ferrand & Branco 2007), and nuclear sequence data (Branco *et al.* 2002; Geraldes *et al.* 2008; Carneiro *et al.* 2013), but no clinal pattern for microsatellites (Queney *et al.* 2001).

The cytonuclear structure in Iberian honey bees is noteworthy given that in Spain, over one million colonies, representing ~50% of existing colonies, have been yearly involved in wide-range movements, in the last decades (A. G. Pajuelo, personal communication). The fact that a marked clinal pattern in both the nuclear and mitochondrial genomes still persists indicates that human-mediated movements play a minor role in shaping Iberian honey bee genetic structure. Nonmutually exclusive explanations can be accounted for the observed pattern. Either transhumance takes place after the reproductive season, or some kind of reproductive barrier or local adaptation is preventing gene flow and long-term establishment of translocated colonies.

Concluding remarks

In this study, a well-defined northeastern–southwestern clinal pattern, revealed simultaneously by nuclear and

maternal markers, provided support for the hypothesis of secondary contact proposed by earlier mtDNA studies. This finding, together with putative signatures of selection detected in a previous study (Chávez-Galarza *et al.* 2013), suggests a complex interplay between adaptation and demography in shaping the Iberian honey bee patterns that we see today. Contemporary human-mediated processes do not seem to be dramatically changing these patterns, a scenario that might change if Spanish and Portuguese beekeepers adopt a strategy of using commercial C-lineage strains, as is occurring in several countries of western Europe (reviewed by De la Rúa *et al.* 2009; Pinto *et al.* 2014). Iberian honey bees are providers of important environmental services through pollination and are number one honey producers in the European Union (European Commission 2013). More importantly, Iberian honey bees represent an important reservoir of diversity that not long ago colonized a broad territory in western Europe (Franck *et al.* 1998; Garnery *et al.* 1998a; Miguel *et al.* 2007). Understanding patterns and underlying processes shaping Iberian honey bee's diversity is an important first step towards preserving this subspecies and thereby the species *Apis mellifera*, an effort of unquestionable value as we face a worldwide honey bee crisis.

Acknowledgements

We are deeply indebted to numerous beekeepers and technicians of beekeepers' associations and of government institutions for inestimable assistance with the honey bee collection. Special appreciation is due to Antonio G. Pajuelo for providing numerous contacts of beekeepers in Spain and answering all the questions related with beekeeping in Spain; to Irene Muñoz, Andreia Brandão, Inês Moura and Margarida Neto for helping in the field work; to Pilar de la Rúa and Irene Muñoz for helping contacting beekeepers in Spain; to Danny Weaver for permission to use the OPA; to Colette Abbey for DNA extractions and running GoldenGate assays; to Clare Gill for assistance with the GenomeStudio software, and Penny Riggs for providing the facilities at Texas A&M University for SNP genotyping; to João C. Azevedo, Miguel Vaz-Pinto and João Paulo Castro for GIS and GPS assistance; to Helena M. Ferreira for searching genomic information of the SNP pairs in LD; and to Liz and Graham Derryberry for providing a script for obtaining the likelihood profiles. GoldenGate SNP data were produced with support from the Texas A&M Institute for Genome Sciences and Society at Texas A&M University in cooperation with Texas A&M AgriLife Research. An earlier version of the manuscript was improved by the constructive comments made by the editor N. Barton and three anonymous reviewers. Julio Chávez-Galarza and Dora Henriques were supported by PhD scholarships from the Fundação para a Ciência e Tecnologia (FCT) (SFRH/BD/68682/2010 and SFRH/BD/84195/2012, respectively). Miguel Carneiro was supported by the same institution through FCT Investigator (IF/00283/2014) and post-doc grants (SFRH/BPD/72343/2010). Financial support for this research was provided by FCT and COMPETE/QREN/EU through the project PTDC/BIA-BEC/099640/2008 to M. A. Pinto.

References

Arias MC, Sheppard WS (1996) Molecular phylogenetics of honey bee subspecies (*Apis mellifera* L.) inferred from mitochondrial DNA sequence. *Molecular Phylogenetics and Evolution*, **5**, 557–566.

Arias MC, Rinderer TE, Sheppard WS (2006) Further characterization of honey bees from the Iberian Peninsula by allozyme, morphometric and mtDNA haplotype analyses. *Journal of Apicultural Research*, **45**, 188–196.

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37–48.

Barton NH, Gale KS (1993) Genetic analysis of hybrid zones. In: *Hybrid Zones and the Evolutionary Process* (ed. Harrison RG), pp. 13–42. Oxford University Press, Oxford.

Barton NH, Hewitt GM (1981) The genetic basis of hybrid inviability in the grasshopper *Podisma pedestris*. *Heredity*, **47**, 367–383.

Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, **16**, 113–148.

Barton NH, Hewitt GM (1989) Adaptation, speciation and hybrid zones. *Nature*, **341**, 497–503.

Beye M, Gattermeier I, Hasselmann M *et al.* (2006) Exceptionally high levels of recombination across the honey bee genome. *Genome Research*, **16**, 1339–1344.

Branco M, Ferrand N, Monnerot M (2000) Phylogeography of the European rabbit (*Oryctolagus cuniculus*) in the Iberian Peninsula inferred from RFLP analysis of the cytochrome b gene. *Heredity*, **85**, 307–317.

Branco M, Monnerot M, Ferrand N, Templeton AR (2002) Postglacial dispersal of the European rabbit (*Oryctolagus cuniculus*) on the Iberian peninsula reconstructed from nested clade and mismatch analyses of mitochondrial DNA genetic variation. *Evolution*, **56**, 792–803.

Campana MG, Hunt HV, Jones H, White J (2011) CorrSieve: software for summarizing and evaluating Structure output. *Molecular Ecology Resources*, **11**, 349–352.

Campos R, Branco M, Weiss S, Ferrand N (2007) Patterns of hemoglobin polymorphism (α-globin (HBA) and β-globin (HBB)) across the contact zone of two distinct phylogeographic lineages of the European rabbit (*Oryctolagus cuniculus*). In: *Phylogeography of Southern European Refugia* (eds Weiss S, Ferrand N), pp. 237–255. Springer, The Netherlands.

Cánovas F, De la Rúa P, Serrano J, Galián J (2008) Geographical patterns of mitochondrial DNA variation in *Apis mellifera iberiensis* (Hymenoptera: Apidae). *Journal of Zoological Systematics and Evolutionary Research*, **46**, 24–30.

Cánovas F, De la Rúa P, Serrano J, Galián J (2011) Microsatellite variability reveals beekeeping influences on Iberian honeybee populations. *Apidologie*, **42**, 235–251.

Carneiro M, Stuart JEB, Afonso S *et al.* (2013) Steep clines within a highly permeable genome across a hybrid zone between two subspecies of the European rabbit. *Molecular Ecology*, **22**, 2511–2525.

Carranza S, Arnold EN, Wade E, Fahd S (2004) Phylogeography of the false smooth snakes, *Macroprotodon* (Serpentes,

Colubridae): mitochondrial DNA sequences show European populations arrived recently from Northwest Africa. *Molecular Phylogenetics and Evolution*, **33**, 523–532.

Chávez-Galarza J, Henriques D, Johnston JS *et al.* (2013) Signatures of selection in the Iberian honey bee (*Apis mellifera iberiensis*) revealed by a genome scan analysis of single nucleotide polymorphisms. *Molecular Ecology*, **22**, 5890–5907.

Chen C, Durand E, Forbes F, François O (2007) Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Molecular Ecology Notes*, **7**, 747–756.

Cornuet JM, Fresnaye J (1989) Biometrical study of honey bee populations from Spain and Portugal. *Apidologie*, **20**, 93–101.

Cosson JF, Hutterer R, Libois R *et al.* (2005) Phylogeographical footprints of the Strait of Gibraltar and Quaternary climatic fluctuations in the western Mediterranean: a case study with the greater white-toothed shrew, *Crocidura russula* (Mammalia: Soricidae). *Molecular Ecology*, **14**, 1151–1162.

Crane E (1999) *The World History of Beekeeping and Honey Hunting*. Routledge, New York.

De la Rúa P, Galián J, Serrano J, Moritz RFA (2001) Genetic structure and distinctness of *Apis mellifera* L. populations from the Canary Islands. *Molecular Ecology*, **10**, 1733–1742.

De la Rúa P, Galián J, Serrano J, Moritz RFA (2002) Microsatellite analysis of non-migratory colonies of *Apis mellifera iberica* from south-eastern Spain. *Journal of Zoological Systematics and Evolutionary Research*, **40**, 164–168.

De la Rúa P, Galián J, Serrano J, Moritz RFA (2003) Genetic structure of Balearic honeybee populations based on microsatellite polymorphism. *Genetics Selection Evolution*, **35**, 339–350.

De la Rúa P, Hernandez-Garcia R, Jiménez Y, Galián J, Serrano J (2005) Biodiversity of *Apis mellifera iberica* (Hymenoptera: Apidae) from northeastern Spain assessed by mitochondrial analysis. *Insect Systematics & Evolution*, **36**, 21–28.

De la Rúa P, Jaffé R, Dall'Olio R, Muñoz I, Serrano J (2009) Biodiversity, conservation and current threats to European honeybees. *Apidologie*, **40**, 263–284.

De la Rúa P, Jiménez Y, Galián J, Serrano J (2004) Evaluation of the biodiversity of honey bee (*Apis mellifera*) populations from eastern Spain. *Journal of Apicultural Research*, **43**, 162–166.

Derryberry EP, Derryberry G, Maley J, Brumfield RT (2014) HZAR: hybrid zone analysis using an R software package. *Molecular Ecology Resources*, **14**, 652–663.

Durrett R, Buttel L, Harrison R (2000) Spatial models for hybrid zones. *Heredity*, **84**, 9–19.

Earl DA, Von Holdt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Endler JA (1977) *Geographic Variation, Speciation, and Clines*. Princeton University Press, Princeton, NJ.

Engel MS (1999) The taxonomy of recent and fossil honey bees (Hymenoptera : Apidae; *Apis*). *Journal of Hymenoptera Research*, **8**, 165–196.

European Commission (2013) EU Market Situation for Honey (ed. Agriculture and Rural Development), pp. 1–20.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Ferrand N, Branco M (2007) The evolutionary history of the European rabbit (*Oryctolagus cuniculus*): major patterns of population differentiation and geographic expansion inferred from protein polymorphism. In: *Phylogeography of Southern European Refugia* (eds Weiss S, Ferrand N), pp. 207–235. Springer, The Netherlands.

Franck P, Garnery L, Celebrano G, Solignac M, Cornuet JM (2000) Hybrid origins of honey bees from italy (*Apis mellifera ligustica*) and sicily (*A. m. sicula*). *Molecular Ecology*, **9**, 907–921.

Franck P, Garnery L, Loiseau A *et al.* (2001) Genetic diversity of the honey bee in Africa: microsatellite and mitochondrial data. *Heredity*, **86**, 420–430.

Franck P, Garnery L, Solignac M, Cornuet JM (1998) The origin of west European subspecies of honeybees (*Apis mellifera*): new insights from microsatellite and mitochondrial data. *Evolution*, **52**, 1119–1134.

Garnery L, Cornuet JM, Solignac M (1992) Evolutionary history of the honey bee *Apis mellifera* inferred from mitochondrial DNA analysis. *Molecular Ecology*, **1**, 145–154.

Garnery L, Franck P, Baudry E *et al.* (1998a) Genetic diversity of the west European honey bee (*Apis mellifera mellifera* and *A. m. iberica*). I. Mitochondrial DNA. *Genetics Selection Evolution*, **30**, S31–S47.

Garnery L, Franck P, Baudry E *et al.* (1998b) Genetic diversity of the west European honey bee (*Apis mellifera mellifera* and *A. m. iberica*). II. Microsatellite loci. *Genetics Selection Evolution*, **30**, S49–S74.

Garnery L, Mosshine EH, Oldroyd BP, Cornuet JM (1995) Mitochondrial DNA variation in Moroccan and Spanish honey bee populations. *Molecular Ecology*, **4**, 465–471.

Garnery L, Solignac M, Celebrano G, Cornuet JM (1993) A simple test using restricted PCR-amplified mitochondrial DNA to study the genetic structure of *Apis mellifera* L. *Experientia*, **49**, 1016–1021.

Gay L, Crochet P-A, Bell DA, Lenormand T (2008) Comparing clines on molecular and phenotypic traits in hybrid zones: a window on tension zone models. *Evolution*, **62**, 2789–2806.

Geraldes A, Carneiro M, Delibes-Mateos M, Villafuerte R, Nachman MW, Ferrand N (2008) Reduced introgression of the Y chromosome between subspecies of the European rabbit (*Oryctolagus cuniculus*) in the Iberian Peninsula. *Molecular Ecology*, **17**, 4489–4499.

Godinho R, Crespo EG, Ferrand N (2008) The limits of mtDNA phylogeography: complex patterns of population history in a highly structured Iberian lizard are only revealed by the use of nuclear markers. *Molecular Ecology*, **17**, 4670–4683.

Gómez A, Lunt DH (2007) Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. In: *Phylogeography of Southern European Refugia* (eds Weiss S, Ferrand N), pp. 155–188. Springer, The Netherlands.

Gonçalves H, Marínez-Solano I, Pereira RJ, Carvalho B, García-París M, Ferrand N (2009) High levels of population subdivision in a morphologically conserved Mediterranean toad (*Alytes cisternasii*) result from recent, multiple refugia: evidence from mtDNA, microsatellites and nuclear genealogies. *Molecular Ecology*, **18**, 5143–5160.

Gowen FC, Maley JM, Cicero C *et al.* (2014) Speciation in Western Scrub-Jays, Haldane's rule, and genetic clines in secondary contact. *BMC Evolutionary Biology*, **14**, 135.

Guillaumet A, Pons JM, Godelle B, Crochet PA (2006) History of the Crested Lark in the Mediterranean region as revealed

by mtDNA sequences and morphology. *Molecular Phylogenetics and Evolution*, **39**, 645–656.

Han F, Wallberg A, Webster MT (2012) From where did the Western honeybee (*Apis mellifera*) originate? *Ecology and Evolution*, **2**, 1949–1957.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. *TAG. Theoretical and Applied Genetics*, **38**, 226–231.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.

Jensen AB, Palmer KA, Boomsma JJ, Pedersen BV (2005) Varying degrees of *Apis mellifera ligustica* introgression in protected populations of the black honeybee, *Apis mellifera mellifera*, in northwest Europe. *Molecular Ecology*, **14**, 93–106.

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, **11**, 94.

Jombart T, Devillard S, Dufour AB, Pontier D (2008) Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity*, **101**, 92–103.

Kandemir I, Kence M, Sheppard WS, Kence A (2006) Mitochondrial DNA variation in honey bee (*Apis mellifera* L.) populations from Turkey. *Journal of Apicultural Research*, **45**, 33–38.

Klaassen CH, Gibbons JG, Fedorova ND, Meis JF, Rokas A (2012) Evidence for genetic differentiation and variable recombination rates among Dutch populations of the opportunistic human pathogen *Aspergillus fumigatus*. *Molecular Ecology*, **21**, 57–70.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews. Genetics*, **4**, 981–994.

Mann HB, Whitney DR (1947) On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, **18**, 59–60.

Meixner MD, Leta MA, Koeniger N, Fuchs S (2011) The honey bees of Ethiopia represent a new subspecies of *Apis mellifera*-*Apis mellifera simensis* n. ssp. *Apidologie*, **42**, 425–437.

Miguel I, Baylac M, Iriondo M *et al.* (2011) Both geometric morphometric and microsatellite data consistently support the differentiation of the *Apis mellifera* M evolutionary branch. *Apidologie*, **42**, 150–161.

Miguel I, Iriondo M, Garnery L, Sheppard WS, Estonba A (2007) Gene flow within the M evolutionary lineage of *Apis mellifera*: role of the Pyrenees, isolation by distance and postglacial re-colonization routes in the western Europe. *Apidologie*, **38**, 141–155.

Miraldo A, Hewitt GM, Paulo OS, Emerson BC (2011) Phylogeography and demographic history of *Lacerta lepida* in the Iberian Peninsula: multiple refugia, range expansions and secondary contact zones. *BMC Evolutionary Biology*, **11**, 170.

Nauta MJ, Weissing FJ (1996) Constraints on allele size at microsatellite loci: implications for genetic differentiation. *Genetics*, **143**, 1021–1032.

Nielsen D, Page RE Jr, Crosland MW (1994) Clinal variation and selection of MDH allozymes in honey bee populations. *Experientia*, **50**, 867–871.

Oleksa A, Chybicki I, Tofilski A, Burczyk J (2011) Nuclear and mitochondrial patterns of introgression into native dark bees (*Apis mellifera mellifera*) in Poland. *Journal of Apicultural Research*, **50**, 116–129.

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics*, **28**, 2537–2539.

Phillips BL, Baird SJ, Moritz C (2004) When vicars meet: a narrow contact zone between morphologically cryptic phylogeographic lineages of the rainforest skink, *Carlia rubrigularis*. *Evolution*, **58**, 1536–1548.

Pinho C, Kaliontzopoulou A, Carretero MA, Harris DJ, Ferrand N (2009) Genetic admixture between the Iberian endemic lizards *Podarcis bocagei* and *Podarcis carbonelli*: evidence for limited natural hybridization and a bimodal hybrid zone. *Journal of Zoological Systematics and Evolutionary Research*, **47**, 368–377.

Pinto MA, Henriques D, Chávez-Galarza J *et al.* (2014) Genetic integrity of the Dark European honey bee (*Apis mellifera mellifera*) from protected populations: a genome-wide assessment using SNPs and mtDNA sequence data. *Journal of Apicultural Research*, **53**, 269–278.

Pinto MA, Henriques D, Neto M *et al.* (2013) Maternal diversity patterns of Ibero-Atlantic populations reveal further complexity of Iberian honey bees. *Apidologie*, **44**, 430–439.

Pinto MA, Muñoz I, Chávez-Galarza J, De la Rúa P (2012) The Atlantic side of the Iberian Peninsula: a hot-spot of novel African honey bee maternal diversity. *Apidologie*, **43**, 663–673.

Pinto MA, Rubink WL, Patton JC, Coulson RN, Johnston JS (2005) Africanization in the United States: replacement of Feral European Honeybees (*Apis mellifera* L.) by an African Hybrid Swarm. *Genetics*, **170**, 1653–1665.

Polechova J, Barton N (2011) Genetic drift widens the expected cline but narrows the expected cline width. *Genetics*, **189**, 227–235.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Queney G, Ferrand N, Weiss S, Mougel F, Monnerot M (2001) Stationary distributions of microsatellite loci between divergent population groups of the European rabbit (*Oryctolagus cuniculus*). *Molecular Biology and Evolution*, **18**, 2169–2178.

R Development Core Team (2013) *R: A language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rortais A, Arnold G, Alburaki M, Legout H, Garnery L (2011) Review of the *Dra*I COI-COII test for the conservation of the black honeybee (*Apis mellifera mellifera*). *Conservation Genetics Resources*, **3**, 383–391.

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.

Rosenberg NA, Mahajan S, Ramachandran S *et al.* (2005) Clines, clusters, and the effect of study design on the inference of human population structure. *PLoS Genetics*, **1**, e70.

Ruttner F (1988) *Biogeography and Taxonomy of Honey Bees*. Springer, Berlin.
Ruttner F, Tassencourt L, Louveaux J (1978) Biometrical-statistical analysis of the geographic variability of *Apis mellifera* L. *Apidologie*, **9**, 363–381.
Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning*, 2nd edn. Cold Spring Harbor Laboratory Press, New York.
Serre D, Pääbo S (2004) Evidence for gradients of human genetic diversity within and among continents. *Genome Research*, **14**, 1679–1685.
Shaibi T, Muñoz I, Dall'Olio R *et al.* (2009) *Apis mellifera* evolutionary lineages in Northern Africa: Libya, where orient meets occident. *Insectes Sociaux*, **56**, 293–300.
Sheppard WS, Meixner MD (2003) *Apis mellifera pomonella*, a new honey bee subspecies from Central Asia. *Apidologie*, **34**, 367–375.
Smith DR, Glenn TC (1995) Allozyme polymorphisms in Spanish honeybees (*Apis mellifera iberica*). *Journal of Heredity*, **86**, 12–16.
Smith DR, Palopoli MF, Taylor BR *et al.* (1991) Geographical overlap of two mitochondrial genomes in Spanish honeybees (*Apis mellifera iberica*). *Journal of Heredity*, **82**, 96–100.
Soland-Reckeweg G, Heckel G, Neumann P, Fluri P, Excoffier L (2009) Gene flow in admixed populations and implications for the conservation of the Western honeybee, *Apis mellifera*. *Journal of Insect Conservation*, **13**, 317–328.
Szymura JM, Barton NH (1986) Genetic analysis of a hybrid zone between the fire-bellied toads, *Bombina bombina* and *B. variegata*, near Cracow in southern Poland. *Evolution*, **40**, 1141–1159.
Szymura JM, Barton NH (1991) The genetic structure of the hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*: comparisons between transects and between loci. *Evolution*, **45**, 237–261.
Tamura K, Peterson D, Peterson N *et al.* (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution*, **28**, 2731–2739.
Yuri T, Jernigan RW, Brumfield RT, Bhagabati NK, Braun MJ (2009) The effect of marker choice on estimated levels of introgression across an avian (Pipridae: *Manacus*) hybrid zone. *Molecular Ecology*, **18**, 4888–4903.
Wallberg A, Han F, Wellhagen G *et al.* (2014) A worldwide survey of genome sequence variation provides insight into the evolutionary history of the honeybee *Apis mellifera*. *Nature Genetics*, **46**, 1081–1088.
Waples RS, Gaggiotti O (2006) Invited review: What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, **15**, 1419–1439.
Weir B (1996) *Genetic Data Analysis*. Sinauer Associates, Sunderland.
Weiss S, Ferrand N (2007) Current perspectives in phylogeography and the significance of South European refugia in the creation and maintenance of European biodiversity. In: *Phylogeography of Southern European Refugia* (eds Weiss S, Ferrand N), pp. 237–255. Springer, The Netherlands.
Whitfield CW, Behura SK, Berlocher SH *et al.* (2006) Thrice out of Africa: ancient and recent expansions of the honey bee, *Apis mellifera*. *Science*, **314**, 642–645.
Wilcoxon F (1945) Individual comparisons by ranking methods. *Biometrics Bulletin*, **1**, 80–83.
Winston M (1987) *The biology of the Honey Bee*. Harvard University Press, London.

M.A.P., J.C.P. and J.S.J. conceived the study. J.C.-G. performed all the analyses. D.H. assisted J.C.-G. with preparation of SNP files and STRUCTURE analyses. M.C. assisted J.C.-G. with cline analyses. M.A.P. coordinated sampling across Iberia and performed sampling across the Atlantic transect. J.S.J. coordinated SNP genotyping. M.A.P. validated SNP genotypes. D.H. assisted M.A.P. with mtDNA PCRs. J.C.-G. analysed the mtDNA sequences. M.A.P. and J.C.-G. interpreted the results with input from J.S.J., M.C. and D.H. J.R. provided the computing environment (software and hardware) for the simulations and analyses. M.A.P. and J.C.-G. wrote the manuscript with input from J.S.J. and M.C.

Data accessibility

SNP genotypes for the 711 Iberian honey bee individuals ordered by sampling location, from AT1 to MT6, in GenePop format: DRYAD entry doi 10.5061/dryad.1kk2k.

Geographic coordinates of sampling locations and environmental data for the 711 Iberian honey bee individuals ordered by sampling location, from AT1 to MT6, in Excel format: DRYAD entry doi 10.5061/dryad.1kk2k.

Aligned mtDNA sequences for 652 Iberian honey bee individuals ordered by sampling location, from AT1 to MT6, in Fasta format: DRYAD entry doi 10.5061/dryad.21s3t.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Statistics of physical distances (bp) of the 309 neutral SNPs used in the genetic analysis of the Iberian honey bee.

Table S2 Pairwise Φ_{PT} values among sampling sites estimated from the neutral SNP data set with GENALEX 6.5 (Peakall & Smouse 2012). Significance of φ_{PT} estimates was assessed using 10,000 permutations. Global Φ_{PT} value was 0.020 (*P*-value=0.001). Pairwise Φ_{PT} values ranged from 0.000 to 0.046. Φ_{PT} values marked in bold were significantly different from zero following Bonferroni correction. Sampling site codes are specified in Fig. 1.

Table S3 Cline parameter estimates for the best-fitting model of 33 SNP loci, complete SNP data set, neutral SNP data set and mtDNA. Cline width is presented as 1/maximum slope. Cline centre and width are measured in km, *P*min and *P*max are the allele frequencies at the ends of the cline, and δ and τ are the shape parameters for the mirror (M), left (L) and right (R) tails. Two log-likelihood unit support limits are presented in parentheses. The symbol * indicates coincident centre and the symbol $^{+}$ indicates concordant width, based on LRTs (*P*-value > 0.05). The left side AICc corresponds to the best-fitting model, and the right side AICc corresponds to the null model. Neutral SNP loci are marked in bold.

Table S4 Linkage disequilibrium (LD) and unbiased haploid genetic diversity (*uh*) estimated from the 309 neutral SNPs for the 23 sampling sites in the Iberian Peninsula (see Fig. 1 for location of sampling sites), as measured by r^2 and percentage of pairs of loci showing significant LD with Fisher's exact test [P_{LD} (Fisher)] before Bonferroni correction (only a single pair remained significant after Bonferroni correction). None of the SNP pairs exhibiting significant LD before Bonferroni correction, for which there is genomic information, are physically linked as they are located in different chromosomes.

Fig. S1 Estimated population structure of *A. m. iberiensis* inferred from the neutral SNP data set (309 loci) by DAPC (top) and STRUCTURE (bottom) at $K = 3$ to 5 clusters. Each of the 711 individuals included in the analyses is represented by a vertical bar partitioned into coloured segments, the size of each corresponding to the individuals' estimated membership proportions in each of the *K* clusters. Black lines separate individuals from the 23 sampling sites, which are arranged from north (AT1, CT1, MT1) to south (AT8, CT9, MT6) in each of the three transects (AT – Atlantic, CT – central, MT – Mediterranean), as indicated at the top bar. Black lines separate individuals from the 23 sampling sites, which are arranged from high *Q* (left) to low *Q* (right) in the blue cluster.

Fig. S2 Graphical display of the three methods (Δ*K*, ΔF_{st} and BIC) to predict the optimal *K* for the analysis of *A. m. iberiensis* population structure using the neutral SNP data set (309 loci) and the complete SNP data set (309 neutral plus 74 putatively selected loci identified by Chávez-Galarza *et al.* 2013).

Fig. S3 Heat map of pairwise Φ_{PT} values between Iberian sampling sites estimated from the neutral SNP data set (309 loci) using GENALEX 6.5 (Peakall & Smouse 2012). The heat map clearly highlights northeastern populations (CT1-3, MT1-2) and CT8 as the most differentiated across Iberia. (a) The 23 sampling sites are arranged from north to south in each of the three transects (see Fig. 1). (b) The 23 sampling sites are arranged along the west–east transect traced for the geographic cline analysis (see dashed line in Fig. 6).

Fig. S4 Population structure of *A. m. iberiensis* estimated by (a) DAPC and (b) STRUCTURE from the neutral SNP data set (309 loci) at $K = 2$ clusters. The 23 sampling sites are arranged along the west–east transect traced for the geographic cline analysis (see dashed line in Fig. 6). Plots represent each of the 711 individuals by a vertical bar partitioned into two coloured segments (blue and red) corresponding to membership proportions (*Q*) in each of the two clusters. Black lines separate individuals from the 23 sampling sites, which are arranged from high *Q* (left) to low *Q* (right) in the blue cluster.

Fig. S5 Mean membership proportion (±SE) in the blue cluster inferred from the neutral (309 loci) and the complete SNP data set (309 neutral plus 74 putatively selected loci identified by Chávez-Galarza *et al.* 2013) with DAPC for each sampling site. Sampling sites are arranged from north (AT1, CT1, MT1) to south (AT8, CT9, MT6) in each of the three transects (AT – Atlantic, CT – central, MT – Mediterranean).

Fig. S6 Maternal pattern, obtained from the $tRNA^{leu}$-cox2 intergenic mitochondrial region, and estimated structure inferred from the neutral SNP data set (309 loci) with DAPC at $K = 2$ clusters. Patterns of variation are displayed at both individual and sampling site level for each transect (AT – Atlantic, CT – central, MT – Mediterranean). Vertical plots display the mitotype (A in red; M in blue; C in orange, one single individual in AT3) and the membership proportions (*Q*) for each of the 711 individuals. Horizontal bar plots show mitotype frequencies (top) and the mean *Q* in blue and red clusters (bottom) at each sampling site.

Fig. S7 Estimated population structure of *A. m. iberiensis* inferred by the spatially explicit algorithm implemented by TESS for the neutral SNP data set (309 loci). Each of the 711 individuals included in the analyses is represented by a vertical bar partitioned into coloured segments, the size of each corresponding to the individuals' estimated *Q* in each of the *K* (from 2 to 5) clusters. Vertical black lines separate individuals from the 23 sampling sites, which are arranged by M (blue), C (orange, one single individual in AT3), and A (red) maternal lineages indicated by colour at the bottom. Sampling sites are arranged from north to south in each transect (AT – Atlantic, CT – central, MT – Mediterranean). Sampling site codes (from AT1 to MT6) are shown at the top bar.

Fig. S8 Plot of DIC values (*Y*-axis) against K_{max} (*X*-axis) obtained with TESS analysis under the admixture model BYM for **(a)** the neutral SNP data set and **(b)** the complete SNP data set.

Fig. S9 Analysis of global and local genetic structure, among 711 individuals of *A. m. iberiensis* from 23 sampling sites, by spatial principal component analysis (sPCA) using the complete SNP data set (309 neutral plus 74 putatively selected loci identified by Chávez-Galarza *et al.* 2013). Each square represents the score of an individual, which is positioned by its spatial coordinates. (a-d) The first four global scores of sPCA. (e-f) The first two local scores of sPCA.

Fig. S10 Spatially explicit analysis implemented by the software TESS for the 711 individuals of *A. m. iberiensis* using the complete SNP data set (309 neutral plus 74 putatively selected loci identified by Chávez-Galarza *et al.* 2013). (a) Plot of individuals' *Q* at the optimal $K = 3$ clusters. Each of the 711 individuals included in the analysis is represented by a vertical bar partitioned into three coloured segments (blue, red, and white) corresponding to *Q* in each of the three clusters. Maternal data (M lineage in blue, C lineage in orange, and A lineage in red) are shown at the bottom. Sampling sites and individuals within sampling sites are arranged as in Fig. 3. (b) Map of the Iberian Peninsula showing the two major clusters ($Q \geq 0.5$)

interpolated by TESS. Dots represent the locations of sampled apiaries across the Atlantic (AT1-8), central (CT1-9) and Mediterranean (MT1-6) transects.

Fig. S11 Map of the Iberian Peninsula with pie charts summarizing frequency data for each sampling site and plot of maximum-likelihood geographic cline for neutral (marked in bold) and selected SNP loci. The symbols * and $^{+}$ indicate the loci or data sets with coincident centre and concordant width, respectively (see Table S3). The dashed line placed in each map represents the transect traced from Lisbon (0 km) to Girona (1400 km) for the geographic cline analysis.